

PN	W09957247-A1
XX	
PD	11-NOV-1999.
XX	
PF	03-MAY-1999;
XX	
PR	01-MAY-1998;
PR	22-OCT-1998;
XX	

PA (REGC) UNIV CALIFORNIA.

aa Fischer RL, Ohad N, Kiyosue T, Yadegari R, Margossian L, Harada J
pi
pi Goldberg RB;

DR WPI: 2000-023577/02.

DR P-PSDB; AAY57037.

XX
XX
XX

used to modulate endosperm development -

XX
PS Claim 11; Page 35; 54pp; English.

This is the FIE3 (fertilisation-independent endosperm) polynucleotide sequence. The FIE genes are a new group of plant regulatory genes. The polypeptide encoded by the nucleotide sequence is involved in the control of reproductive development. The FIE sequence shares homology with a polycomb group gene from *Drosophila*. The invention relates to FIE1 and FIE3 polynucleotide sequences (AA37895-237896). The nucleotide sequences can be used to create a transgenic plant containing an expression cassette with an FIE nucleotide sequence operably linked to a plant promoter. The FIE polynucleotide can be used to modulate endosperm development in a plant. The nucleotide sequences can be used to increase or enhance endogenous gene expression e.g. to enhance the expression of polycomb genes which control cell cycling. This can be used to control plant morphology. Enhanced expression can also be used to increase vegetative growth.

Sequence 1563 BP; 429 A; 272 C; 368 G; 494 T; 0 other;

Query Match	100.0%	Score	1563	DB	21	Length	1563
Best Local Similarity	100.0%	Pred. No.	0				
Matches 1563; Conservative	0	Mismatches	0			Indels	0
						Gaps	0

Qy	1	aaagtgagctgtgtgtgtgtgtgtcaggtcccaataaagtttgcgttggtgtccaaatct	60
Db	1	aaagtgagctgtgtgtgtgtgtgtcaggtcccaataaagtttgcgttggtgtccaaatct	60
Qy	61	aaggttaacgaatatttaataactctggaatctgtgtcctaactgaataatacaaacacc	120
Db	61	aaggttaacgaatatttaataactctggaatctgtgtcctaactgaataatacaaacacc	120
Qy	121	agttgtgtgtcatalpagaagacttgagctttagatlttagtvtgtcggagagtcacagaagag	180
Db	121	agttgtgtgtcatalpagaagacttgagctttagatlttagtvtgtcggagagtcacagaagag	180
Qy	181	agagattccgaataatcgaaatgtcgaaagataaacttaaggagacgaatgaatagtgtgtct	240
Db	181	agagattccgaataatcgaaatgtcgaaagataaacttaaggagacgaatgaatagaattgtgtct	240
Qy	241	ttagactccaatgaataagaataatcgtaacaaatgtacgaaatgaaatcagaaggaagaa	300
Db	241	ttagactccaatgaataagaataatcgtaacaaatgtacgaaatgaaatcagaaggaagaa	300
Qy	301	cccttgatgctgtgtgttttcaacttccctgaatgtccglttctcgatgcttcgttacc	360
Db	301	cccttgatgctgtgtgttttcaacttccctgaatgtccglttctcgatgcttcgttacc	360
Qy	361	gctgtgtggaataccgattactctgacaatattgtctcgaaagtgtgtgcataaagacttg	420
Db	361	gctgtgtggaataccgattactctgacaatattgtctcgaaagtgtgtgcataaagacttg	420
Qy	421	caactctaagctgataagaataagaagaagtcgttttcaacaggtgaagtlttggtcggtgac	480
Db	421	caactctaagctgataagaataagaagaagtcgttttcaacaggtgaagtlttggtcggtgac	480
Qy	481	gttaatggaaccacataatgtgtggtcgtggaagagttaaaggtataatccgaatcatggac	540
Db	481	gttaatggaaccacataatgtgtggtcgtggaagagttaaaggtataatccgaatcatggac	540
Qy	541	gtcaacaagtgaaacgaattataagaagcttlttggttcaatgaagttcaagtgaagaatc	600
Db	541	gtcaacaagtgaaacgaattataagaagcttlttggttcaatgaagttcaagtgaagaatc	600

Oy	601	agagcacaaaccttlaaaacctcaactctgtatctactctgtagcaagagatgaaatcgtttccg	660
Db	601	agggcacaaacctttaaaccctcaactctgtatctactctgtagcaagagatgaaatcgtttccg	660
Oy	661	ttgtgtgaatgttgaactcggggaatgattttgttatcttgccttggaagcgtggaggtcaccgc	720
Db	661	ttgtgtgaatgttgaactcggggaatgattttgttatcttgccttggaagcgtggaggtcaccgc	720
Oy	721	tataaaattcctaagttgtggattcttcacctccgtctgatatattaaccccttctgtactgttgcgt	780
Db	721	tatgaagttctaaagtctgagatcttccatccgtctcgtatatttaacgccttctgtactgttgcgt	780
Oy	781	atggacacccaactatlaaataatggtcgaatgaagaagttttggaagctacgctcggaaagtca	840
Db	781	atggacacccaactatlaaataatggtcgaatgaagaagttttggaagctacgctcggaaagtca	840
Oy	841	ttcacatgagactgtatgccatcaaaaaattcccccaaaaaattgttccaatttccctgtatt	900
Db	841	ttcacatgagactgtatgccatcaaaaaattcccccaaaaaattgttccaatttccctgtatt	900
Oy	901	acagtttccattctatacaaatittgtagattgtgaacgtttgtttgtgtatttatccctc	960
Db	901	acagtttccattctatacaaatittgtagattgtgaacgtttgtttgtgtatttatccctc	960
Oy	961	tcaaaagatgttggaaccaacgagatccgtttgttgggaaaccaacaactgaaagagaattccct	1020
Db	961	tcaaaagatgttggaaccaacgagatccgtttgttgggaaaccaacaactgaaagagaattccct	1020
Oy	1021	ggcgaaggagactcgaatgttctatattaagaataccggttccaattgttgatatttgggtt	1080
Db	1021	ggcgaaggagactcgaatgttctatattaagaataccggttccaattgttgatatttgggtt	1080
Oy	1081	atcaagaattttcttcttgaccctccatttaagaatctcgtttgagatgaatgaatcgaagaagaag	1140
Db	1081	atcaagaattttcttcttgaccctccatttaagaatctcgtttgagatgaatgaatcgaagaagaag	1140
Oy	1141	gtttatgtctcggatttgtaaaagtctgcctccgttttgattatcaaaagtatcacacaat	1200
Db	1141	gtttatgtctcggatttgtaaaagtctgcctccgttttgattatcaaaagtatcacacaat	1200
Oy	1201	caatcaaaagctctgatatcagaagcaaacagcgaatctctgtctgatacgaatgagcagatcttgc	1260
Db	1201	caatcaaaagctctgatatcagaagcaaacagcgaatctctgtctgatacgaatgagcagatcttgc	1260
Oy	1261	tgcctcggagaaaggagactatattgscgtcttggaacgttatataccaagaatgagcgtctgagct	1320
Db	1261	tgcctcggagaaaggagactatattgscgtcttggaacgttatataccaagaatgagcgtctgagct	1320
Oy	1321	tgtagaagaatctgatactagatagagatgcgaagaatagatataccatcttcttattgttaact	1380
Db	1321	tgtagaagaatctgatactagatagagatgcgaagaatagatataccatcttcttattgttaact	1380
Oy	1381	cttgatcatgtgtctactccctgagaaaccttgagatgagctcttctgtgagcctctgtttaacgtcca	1440
Db	1381	cttgatcatgtgtctactccctgagaaaccttgagatgagctcttctgtgagcctctgtttaacgtcca	1440
Oy	1441	ccctctgtaccacaagtgatatcccttctcggagatattgttgcttatcttcttaattttcaaac	1500
Db	1441	ccctctgtaccacaagtgatatcccttctcggagatattgttgcttatcttcttaattttcaaac	1500
Oy	1501	acaaggctgtaccctcggagacttattgtcagagaacaacctctcttctcaataagctttctagtta	1560
Db	1501	acaaggctgtaccctcggagacttattgtcagagaacaacctctcttctcaataagctttctagtta	1560
Oy	1561	ttc 1563	
Db	1561	ttc 1563	
RESULT	2		
AA501160			
AA501160			
standard; cDNA; 1626 BP.			

KW seed; pharmaceutical; nutraceutical; polymer; rice; rcallc.pk0001.d2.
XX
OS Oryza sativa.
FH Key Location/Qualifiers
FT CDS 25..1188 /*tag= a
FT /product= "Fertilisation-independent endosperm protein"
XX
FN WO200116325-A2.
XX
XX 08-MAR-2001.
XX PD
XX PE 30-AUG-2000; 2000WO-US23735.
XX PR 31-AUG-1999; 99US-0151575.
XX PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
PI Simmons CR, Weng Z;
XX WPI: 2001-244407/25.
DR P-PSDB; AAU00326.
XX
XX New plant fertilization independent endosperm protein for the
PT production of seed without fertilization is recombinantly produced -
XX
PS Claim 7; Page 67; 94pp; English.
XX
CC The sequence represents the coding sequence of fertilisation-
CC independent endosperm clone rcallc.pk0001.d2. Fertilisation-independent
CC endosperm proteins are plant reproduction proteins necessary for
CC apomixis, the formation of seeds without fertilisation. Apomixis is
CC especially useful to agriculture because it eliminates the necessity of
CC selfing plants to produce genetically identical seed. Such seed is useful
CC to produce seeds for human and animal food and for commercial milling and
CC extraction, including the production of useful recombinant products in
CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and
CC polymers. Embryoless seed production by transgenic plants is less likely
CC to cause ethical and environmental concern over transgenic plant
CC production as no gametes are being produced to cross pollinate with other
CC crops and the seeds cannot germinate.
XX
SQ Sequence 1387 BP; 389 A; 246 C; 329 G; 422 T; 1 other;

Query Match 38.5%; Score 602; DB 22; Length 1387;
Best Local Similarity 72.3%; Pred. No. 1.7e-167;
Matches 782; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

QY 214 tttaggaacgaatcaatagttggtcttgactcgaatcaagaatcgtacaagt 273
DB 70 ttggagatgagaccagtggtggtatcctgtgttcgaagaagaagaatacagaagtc 129
QY 274 acgaatagatccagaagaagaacacctgtatgctgtgttttcaacttcctgat 333
DB 130 accaatgcctccaagagggaagacgcctctatacgcgcgtatcttcaactcattcgc 189
QY 334 gctgttcttctgtagtctgtaccgctgtgtgaatcgaattactctgtacattgt 393
DB 190 tcccgactactcaagtttcgcacatgttgcggaataaggtactctgttatacattgc 249
QY 394 ctccgagatggtcattacatcgaattgcattcgtcgaagaagaagaagtgc 453
DB 250 ctgtatgaaggagatattgctgtttgcaattatgtagagataagaatgagct 309
QY 454 tttaacaaggtaagttggcgctgtgcgttaattgggaaccataatgtgcgctggagga 513
DB 310 tttaacacggtgtgtggtgcatgcaatgttgaacgggaacccactgttgtgctggagga 369
QY 514 gtaaaaggtatatacgcagatcattgacgtcaacagtgtaaaagattacatagatctgtg 573

DB 370 ctcaatggggtaatccgagtcattgtatgtctgtagcagtgagaagaataataagatttgt 429
QY 574 ggtcatgagatcagtgacaagaatcaggacacaaccttaaaccttaactgtgatt 633
DB 430 ggccatgagactcctaataatgaagccaagccaataataatcactcactgtgta 489
QY 634 actgctcagaagaagtaaatctgttctgtgtgtgaatgtgaactggatgattgtg 693
DB 490 tcgycagaagaagatgaactatctcgttctgtgaatgaaagccatcctggaatgactgtg 549
QY 694 atatttgcctggagctcagtgatccgtatcgaattcgaattcgaattcattccct 753
DB 550 atatttgcctggagctcagtgatccgtatcgaattcgaattcgaattcattccatgc 609
QY 754 gatattaccgctctgtcgtgtgtgtatgacacacactataaataatgtaataagaa 813
DB 610 gatatgtatcgtatctgt 669
QY 814 gaggtttgcagtgatcgtcgaagaatcattcacaatgagctgtgattccataaatcc 873
DB 670 gaggtttgcagtgatcgtcgaagaatcattcacaatgagctgtgattccataaatcc 729
QY 874 acaaaattgtccaatctcctgtatcagctcattcattcacaatcattatgtatgt 933
DB 730 acaaaattgtccaatctcctgtatcagctcattcattcacaatcattatgtatgt 789
QY 934 aacgctgt 993
DB 790 aatagtggt 849
QY 994 gaaccaacatcgaagaatctcctcgtcgaaggagcttcagatgtctataatgatac 1053
DB 850 gaacctaaagtgaaacgaacaactccaggaagagtgatgtgtgtgtgtgtgtgtgtgt 909
QY 1054 ccggttcgaatgtgataattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1113
DB 910 cctatcccgatgtcattcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 969
QY 1114 gttgcgatagtgtaacgaagaagaagtttatgtctgtgatttgaagaagttgcctct 1173
DB 970 gttacagtggtgaacgaagaagaagatttgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1029
QY 1174 gtttgattacaagtattcacacaatcaatacaatcgtgaactcggcaacagcattg 1233
DB 1030 gtaactgtcgaagaagttgtcacatcctcaatcaataatcccaatcaggcagctgca 1089
QY 1234 tctgtcgtatggaagcagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1293
DB 1090 tctcttgatggaagtactatattgagtgctgtgaggaatggacatatgagcgttgg 1149
QY 1294 gt 1295
DB 1150 gt 1151

RESULT 5
AAS01172
ID AAS01172 standard; cDNA; 1498 BP.
XX
XX AAS01172;
XX
XX 12-SEP-2001 (first entry)
XX
XX Fertilisation-independent endosperm cDNA clone p0120.cdeb48r.
KW Fertilisation-independent endosperm; plant reproduction; apomixis; ss;
KW seed; pharmaceutical; nutraceutical; polymer; rice; p0120.cdeb48r.
XX
XX Oryza sativa.
XX
XX Key Location/Qualifiers
FH CDS 3..1130

OY	600	caggaaccaacctttaaactoaacttgatctggattacgtcgtagcaaggataactgctgc	659
Db	413	aagaactccaagcattaaagcccttcgccaccatcatcttcgcaagcaaggatgatctgttag	472
OY	660	tttgtgaatctgtbaaacctggtgatatgtlatcttgatataatttgtctggaagctggatccg	719
Db	473	gctctggaagtgttccacaacaggagatctgcatcttgatttttgtctggaagcaaggtccacg	532
OY	720	ctatgaagcttcaaagtgtggaattttccatccgctgcgtgataatttaacgcctttggttaattg	779
Db	533	gaatgaagtatgtgattgttgacttcccaccatctgfatatctaccgcatagaagaattgttgg	592
OY	780	tatggaacaccactttaaatatgataatgaagaagadttttggacgctaagctcggaagact	839
Db	593	catgataaacctcgtltaaaaaatgltcaatggaagaattctcgcccatatgtttggaacatc	652
OY	840	attcacatgagctatgatccatcaaaaattccccacaaaattttgccaatccctgtaat	899
Db	653	cttcaatagagctgaaccttccatcataaaattccacaanaaatatgtgcaattccggtct	712
OY	900	tacagcttccatgcatacaaaatlattgtagattgttaaccglttggtttggttattatccct	959
Db	713	gcttgctgtagatcatcttaactactatgtgattgtactagatagctgtgtgtacttcatct	772
OY	960	ctcaaaaggtctggaacaacgagatccgtgtgttggaacaccacaactggaaggaattctcc	1019
Db	773	gtcaaaaggtgttgacaatgaattgtgcgtgtggagccaaaacaaaagaacaagaatcc	832
OY	1020	tggcgaaggagcttcacagatgtcttataagaataccggttccaatgtgtgatalttggt	1079
Db	833	cgggaaggtatgcatgtgatatcttccaagaatattccgtgtgcagaagtgtgatalctggt	892
OY	1080	tatcaagtttctgtgacccctccatllaagtctgtgtgcgataaggtlaatcsgaagaa	1139
Db	893	tatcaaatcttcacatgcgatttcaacttcaatccaattggaataagcaacgltgaaggaaa	952
OY	1140	ggttatagtctcgggatttgaagaattgccccttcgttttttgattacaagaattatcacaaa	1199
Db	953	agctctgtctcgggaagatcacagctccagctccctcgttttlaactgtcgcgtgactcaatcc	1012
OY	1200	tcaatcaaaagctgtaatacgaacgaacgcatgctgtcgaatggaagcacagatcttgc	1259
Db	1013	gcaatgcaaatctcgataaaggcaagctgcgtgtcatcttgatggaagcaaatccctgc	1072
OY	1260	ttgtctgcaaggaacggagactatataggcgtggaagc	1294
Db	1073	ctgcagcgagatgcaagcatatgscgatgtggtatg	1107
RESULT 6			
AASO1174			
ID	AASO1174	standard; CDNA; 1518 BP.	
XX			
NC	AASO1174;		
XX			
DT	12-SEP-2001 (first entry)		
XX			
DE	Fertilisation-independent endosperm cDNA clone ses2w.pk0015.b10.		
XX			
KW	Fertilisation-independent endosperm; plant reproduction; apomixis; ss;		
KM	seed; pharmaceutical; nutraceutical; polymer; wheat; ses2w.pk0015.b10.		
OS	Triticum aestivum.		
XX			
FH	Key Location/Qualifiers		
FT	CDS	88..1224	
FT	/*tag= a		
XX	/product= "Fertilisation-independent endosperm protein"		
XX	WO200116325-A2.		
PD	08-MAR-2001.		
XX			

Db	595	atttttctgtgaaggaggaaggtcaacgcgaataatgaagtattgaagtgttgaacttccaccctctc	654
Qy	754	gatatltacccgttctgtctagttgtgtgatgagacacacattaaatatgctgaatgaa	813
Db	655	gatatctacccgaatttcccgattgtgtgtgatgataataatctgttaaatctgtcgaatgaa	714
Qy	814	gagttttggaagtacgtccgggaaggtcattcaacttgaactcgtatgataccaataatcccc	873
Db	715	gaattttgcacatcgtgtgagaataatcctttacatgtgacttgcaccttccatcaaaatttcca	774
Qy	874	acaaatattgcacaaattccctgtatattacacgcttccatccatatacaaatatgataatgtc	933
Db	775	acgaatttgttcaatttcgcgtcatgacttcgcgttgattcattccaactatgttgacgt	834
Qy	934	aacggttggtttgtgtatttatctctccaagaagtgtgtgacaaagagatccctgttctg	993
Db	835	actagttgcttgttgcattctactctgtccgaagagtgattgacaaatgaatttcttctg	894
Qy	994	gaaccacacacgaagaagaattctccctcgggagagggagcttcagatgtttctataagatac	1053
Db	895	gagccaaaaataaaagagcaggggtcccgaggaggtgacatgattgatttccaagaatac	954
Qy	1054	ccggttccaatgtgtgatatttggttatacaagtatttcttctgtgacctccattaaagtctc	1113
Db	955	ccctgtgcctgtatgtgacatttggttatatacaattctcagttgatttccaacttaacaa	1014
Qy	1114	gttgcgtgaggttaatcggaggaagaaggttatatgtcttgcgatttgaagaattggccctct	1173
Db	1015	tttagcaataatgacacccggaggaacaaatctatgtgttgggaagtgcagtcggagccctct	1074
Qy	1174	gtttgttatcaaaagtatacaacaacaaatcaaaagtctgtatccagtgaaacagcagcat	1233
Db	1075	gtgtcctaattaccggctgagtagtgcacaaatgcaaaatgcaataatgcataagcagagctcatg	1134

Db 1135 tcgtttgtagaagcagcatccttgccttcgcgcgaagatgtgcagcatataccgcctgggat 1194

Db	1195	gaagtgaacatcaagctgcataaaattgaagcactga	1233
RESULT 7			
AS01156			
ID	AS01156	standard; cDNA; 1643 bp.	
XX	AS01156;		
AC			
XX			
DT	12-SEP-2001	(first entry)	
XX			
DE	Fertilisation-independent endosperm cDNA clone ccase-b.pk0026.g4.		
XX			
KW	Fertilisation-independent endosperm; plant reproduction; apomixis; ss;		
XX	seed; pharmaceutical; nutraceutical; polymer; ccase-b.pk0026.g4; corn.		
XX			
OS	Zea mays.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	193..1332	
FT	/*tag= a		
FT	/product=	"Fertilisation-independent endosperm protein	
XX			
PN	MO200116325-A2.		
XX			
PD	08-MAR-2001.		
XX			
PF	30-AUG-2000; 2000MO-US23735.		
XX			
PR	31-AUG-1999; 99US-0151575.		
XX			
PA	(DUP0) DU PONT DE NEMOURS & CO E I.		
PA	(PION-) PIONEER HI-BRED INT INC.		

CC The sequence represents the coding sequence of fertilisation-independent endosperm clone csa-*b*.PK0026.g4. Fertilisation-independent endosperm proteins are plant reproduction proteins necessary for apomixis, the formation of seeds without fertilisation. Apomixis is especially useful to agriculture because it eliminates the necessity of selling plants to produce genetically identical seed. Such seed is useful to produce seeds for human and animal food and for commercial milling and extraction, including the production of useful recombinant products in the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and polymers. Embryoless seed production by transgenic plants is less likely to cause ethical and environmental concern over transgenic plant production as no gametes are being produced to cross pollinate with other crops and the seeds cannot germinate.

Sequence 1643 BP; 429 A; 345 C; 456 G; 413 T; 0 other;

Query Match	31.3%	Score 489.4	DB 22	Length 1643
Best Local Similarity	66.4%	Pred No. 3.7e-134		
Matches 719; Conservative	0	Mismatches 361	Indels 3	Gaps 1

QY	215	taaggaaacgaatccaatagtgtgggtccttgtaactccatccatgcgaataaagaatccgtaacaagtga	274
Db	221	tcgggtctggagacgcgacggagaggctgcgtctgcacagccggaagccgggaataacaacccct	280
QY	275	cgaatcagatccaagaagaaggagaacaacctgtatacgtctgtgtttccaacttcctatg	334
Db	281	gcggcaagacacacctgagggaggagccgcgcataatgtatccggttccaacttcatacgacg	340
QY	335	cctcgttcttcgaatgtcttcgttaacgcgtgtgtgnaatcggatctactctgtacaaattgc	394
Db	341	cgcgtactacaagcgtcttcgcacacgtcgcgcgacacacgcggtfacacaacttaccgtgc	400
QY	395	tcggagatggtgtcccatccatccagatctgaatccctctgtctgttgtaagaagaagaagtcgt	454
Db	401	ttgagatgtgtatgttctgcctctccacaacgcttaacgttctgtatgaaagaaatgaatcgt	460
QY	455	tttacaacgttaagttgggcgtgtgc---gttaatgggaacccaatgltgtcgcgtgcgg	511
Db	461	tctatactcttaagcctcgaggctcgtgaaccaatgltgtatgtctcaaccaactcgtgtgacgaag	520
QY	512	gaagtaaaagataatcccgagtcatttgacgtccaacagtgaaacgattccataagaagctctg	571
Db	521	gaagcaatggagtcattccggttcatacaattgtgtacaagaagaagttgacataagaagcttgc	580
QY	572	tgggtcatgtgaattcagtgaaacggaatccggaacaacaacctttaaaccccaactgtga	631
Db	581	ttggccatgtgcactccaataaatatgagataaagaacaacacgcgtgtgaagccttcgtccatca	640
QY	632	ttacgtctgacgaagatgtgatactgtctcgttgttggaatcgttgaaacccggggatatagtatt	691
Db	641	ttctctgaagcaagatgatactcgtttaagcgtatagatgaatgtccatacaaggaatccgtatct	700
QY	692	tgaattcgtcgaagcctggaagtcacgcgtatgaagaagttctaagtgtgaatttccatccgt	751
Db	701	tgtatttgcgtgagcgtggaagtcacgcgaatgaagaatataagtgatgtactccatccacta	760
QY	752	ctgaatatccaacgcttctgtctaattgtgtatgagaacaacacattaaaataatgggtcaatga	811
Db	761	gtatattgaacgttttctgaagttgtgtgatgagaacaacacgtgtgaataatccgtgtcaatga	820

[illegible]

RESULT	8
AAS01170	
ID	AAS01170 standard; cDNA; 1686 BP.
XX	
AC	AAS01170;
XX	
DT	12-SEP-2001 (first entry)
XX	
DE	Fertilisation-independent endosperm cDNA clone p0107.cbcal79r.
XX	
KW	Fertilisation-independent endosperm; plant reproduction; apomixis; ss;
KM	seed; pharmaceutical; nutraceutical; polymer; p0107.cbcal79r; corn.
XX	
OS	Zea mays.
XX	
FH	key
FT	location/Qualifiers
FT	CDS
FT	394..1533
FT	/tag= a
XX	/product= "Fertilisation-independent endosperm protein"
PN	WO200116325-A2.
XX	
PD	08-MAR-2001.
XX	
Pf	30-AUG-2000; 2000MO-US23735.
XX	
PR	31-AUG-1999; 99US-0151575.
XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E I.
PA	(PION-) PIONEER HI-BRED INT INC.
XX	
PI	Bittler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
PI	Slimmons CR, Weng Z;
XX	
WPI	WPI; 2001-244407/25.
DR	P-PSDB; AAU00323.

PS Claim 7; Page 58; 94pp; English
XX

CC The sequence represents the coding sequence of fertilisation-independent endosperm clone p0101.cgmg48r. Fertilisation-independent endosperm proteins are plant reproduction proteins necessary for CC apomixis, the formation of seeds without fertilisation. Apomixis is CC especially useful to agriculture because it eliminates the necessity of CC selling plants to produce genetically identical seed. Such seed is useful CC to produce seeds for human and animal food and for commercial milling and CC extraction, including the production of useful recombinant products in CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and CC polymers. Embryoless seed production by transgenic plants is less likely CC to cause ethical and environmental concern over transgenic plant CC production as no gametes are being produced to cross pollinate with other CC crops and the seeds cannot germinate.

CC Sequence 1803 BP; 446 A; 413 C; 491 G; 449 T; 4 other;

Query Match 31.1%; Score 486.2; DB 22; Length 1803;
Best Local Similarity 66.2%; Pred. No. 3,4e-133;
Matches 717; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

QY 215 taaggacagatcaatagttggtcttgaatcgcatacgaataaagcgtacaagtgta 274
DB 518 tcggggtcgagagcgcgaggggtcgtcgccagccggaagcgaggatacaagccct 577
QY 275 cgaatagatcaggaaggaagaaacattgtatgctgtgtttcaacttcctatg 334
DB 578 gcggcagacacactatggggaagcgccgctatgtatcgtggttcaactcatgagc 637
QY 335 ctgcttctcgatgctcttcgttaacgctggtggaatcgatctactctgtaattgtc 394
DB 638 cgcgctactacagctcttccgacccgtcgcggaacgcggaacactacgcgcgcgc 697
QY 395 tcggagatggtgcatacgcatacgcatacgcatacgcatacgcatacgcatacgc 454
DB 698 ttgagaatgtatatttcgctcttctacaagcttaacgttggtaggaagatgagtcgt 757
QY 455 tttaacgtaagttggcgctgtgc--gttaatgggaacccaatgctggtcgtag 511
DB 758 tctatactcaagctggcgctgacatgtgtatggtcaccactgtgtgtagcagcag 817
QY 512 gagtaaaagatatacccgatcattgaactcaacagtgaaacgatcataagagcttg 571
DB 818 gaagcaatggatcattcgtggtatcattcgtgtacagaagaattgacgaagccttg 877
QY 572 tgggtatgagatcagtgaaacgaatacagacacacacttaaaacctcaactgtga 631
DB 878 ttggcctatggactataaataatgataagaactcaacgtttgaagccttcgcatca 937
QY 632 ttaactgtaagaagatgaatctgtctgtgtggaatgttgaactggaatgtatct 691
DB 938 ttctgcaagaagaatgaaatctgttaggtatgaaatgcatcagaaggtctgtatct 997
QY 692 tgatatctgtgagcggtggaagtcatacgctatgaaatgtaagtgtgattcatcct 751
DB 998 tgatatctgtgagcggtggaagtcatacgctatgaaatgtaagtgtgattcatcct 1057
QY 752 ctgattattacgcttctgtgctgtgtgtatggaacacacattaaatgagtcata 811
DB 1058 gtgataatgaaagcttctgcaagttgtgcatgagcaaaacgttgaacatcggtca 1117
QY 812 aagagtttgcagctgacgagatcattcattcagatgagatgagtcatacaaatcc 871
DB 1118 aagatcttgcataatgttgacaacataatcttcatgagcagcattcattcaagtttc 1177
QY 872 ccacaaatgttccaaatccctgtatttatacagcttccattcattacataatgtaggt 931
DB 1178 caacaaatgttccaaatccctgtatttatacagcttccattcattacataatgtag 1237
QY 932 gtaacgcttgttggtagatttatacctcctaagaagtgtagacaagagatccgtgt 991
DB 1238 gtacaagatggtcgtgtagcttccatacaagaagtgtagacaatgtgtcttt 1297

QY 992 gggaaaccacaactgaaagaatctcctgagcagggagcttcagatgltcattaatg 1051
DB 1298 gggaaaccgaacaaagaacagagctcctgggagggagacatcagataccttcagag 1357
QY 1052 acccggttccaaatgtgtatattgtttataaagttttctgttactccattgaagt 1111
DB 1358 acccggttccaaatgtgtatattgtttataaagttttctgttactccattgaagt 1417
QY 1112 cgtgtgcatgagatcaacgaaggaaggttattgctcgtgatttgaagaattgcctc 1171
DB 1418 agttgcatgagcaacgctggaaggaacaaatcagtggtggaagtacagtcagccctc 1477
QY 1172 cgtgttcttcaaaagtatacacaacatacgaatcgttgaatcaggaacaagcca 1231
DB 1478 ctgtctcattgtcgtcgtgtatatacagcagtgatcagcagataagaacaactgcag 1537
QY 1232 tgtctgcatgaggaacagatcttctgtcgtcgtgaggaagcagactatagcgtg 1291
DB 1538 tgtccttcgatggaagcacaatccttggagctgtgtgaagcagcaccactcgtg 1597
QY 1292 acg 1294
DB 1598 atg 1600

RESULT 10

AA501164
ID AA501164 standard; cDNA; 1486 BP.

AA501164;
AC
XX
XX

DT 12-SEP-2001 (first entry)

DE Fertilisation-independent endosperm cDNA clone p0003.cgped29rb.

XX Fertilisation-independent endosperm; plant reproduction; apomixis; ss;

KW seed; pharmaceutical; nutraceutical; polymer; p0003.cgped29rb; corn;

OS EST; expressed sequence tag.

OS Zea mays.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

W0200116325-A2.

08-MAR-2001.

30-AUG-2000; 2000WO-US23735.

31-AUG-1999; 99US-0151575.

PA (DUPO) DU PONT DE NEMOURS & CO E. I.

PA (PION-) PIONEER HI-BRED INT. INC.

PI Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;

PI Simmons CR, Weng Z;

DR WPI; 2001-244407/25.

DR P-PSDB; AAU00317.

XX New plant fertilization independent endosperm protein for the

PT production of seed without fertilization is recombinantly produced -

XX Claim 7; Page 51-52; 94pp; English.

XX The sequence represents the coding sequence of fertilisation-

XX independent endosperm clone p0003.cgped29rb. Fertilisation-independent

CC endosperm proteins are plant reproduction proteins necessary for
 CC apomixis, the formation of seeds without fertilisation. Apomixis is
 CC especially useful to agriculture because it eliminates the necessity of
 CC selling plants to produce genetically identical seed. Such seed is useful
 CC to produce seeds for human and animal food and for commercial milling and
 CC extraction, including the production of useful recombinant products in
 CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and
 CC polymers. Embryoless seed production by transgenic plants is less likely
 CC to cause ethical and environmental concern over transgenic plant
 CC production as no gametes are being produced to cross pollinate with other
 CC crops and the seeds cannot germinate.

CC Sequence 1486 BP; 391 A; 306 C; 380 G; 408 T; 1 other:

DB 927 tgtgacatttggttatacaatttcatgatttcttcaatcatcagttgagatagc 986
 QY 1126 aatcagaaggaaggttattatgtctggtattgaaagtgtccctcttttgatataa 1185
 DB 987 aaccgtgaagcgaatactacgtgtggaagt-acagtaagcctcttccatcatgtct 1045
 QY 1186 aagttacacatcaatcaaatcgttgaatcaaggaacagccatgctgtcgatgga 1245
 DB 1046 cggctgtataatcagcagtgtaaatccgcgataagaacaatcgtcgtccttgatgga 1105
 QY 1246 agcacgattctgtctgtcgcgagagcggactataatgctgctggagc 1294
 DB 1106 agcacatctcttgagctgtgtaagagcgcacatctgctggtggatg 1154

Query Match 30.2%; Score 472.2; DB 22; Length 1486;
 Best Local Similarity 68.1%; Pred. No. 4.3e-129;
 Matches 687; Conservative 0; Mismatches 318; Indels 4; Gaps 2;

QY 289 gaaggagaagaacaccttgaatctgttcttcaacttctcagatgctgttcttccat 348
 DB 147 gaaggagaagcgcgcgtatagctacggttcaacttgaagcgcgcgtactacagac 206
 QY 349 gcttcgttaccgctgtggaatcagatctgttacaattgtctcggagatgtgccc 408
 DB 207 gcttcgcacccgttcggcggaacccgctgacaacttalcgctgtccttgagatgtgta 266
 QY 409 atatacgaatcgaatcctatgctatgataagaagaagtcgttttaccagtaagt 468
 DB 267 ttcgctcttctatacgaatcgttgcgtatgagataagatgctgttctactactaagc 326
 QY 469 tggggtgtggc---gttaatgggaaccatattgttgcgcctggaggaataaagatga 525
 DB 327 tgggctcgtgacacgttgaatgctcaccatcgttggcgacggagaagcaatggagac 386
 QY 526 atccgagtcacatgacgtcacacagtgaaacgattacataagcttctgtggtacatgag 585
 DB 387 atccggtcatcatattgtgtctacagaagaagttagctaaagcttgttgccatgtgcac 446
 QY 586 tcaatgaacgaatacaggaacacaccttaaaacctgaactgtgatactgctagcaag 645
 DB 447 tcaataaattgataagaactcaacaccttgaagcctgtcatcatcattcttgcagaag 506
 QY 646 gatgaatcgttctgttgggaattgtgaactggaatattgtatttctgtga 705
 DB 507 gatgaatcgttctgttgggaattgtgaactggaatattgtatttctgtga 566
 QY 706 gctggaagtcacatgataaagttctaagttggaatttccatccgctcatatattccgc 765
 DB 567 gctggaagtcacatgataaagttctaagttggaatttccatccgctcatatattccgc 626
 QY 766 ttgtcgaattgtgtatgagacacacatataaataatgtaacgaagaagtttggagc 825
 DB 627 ttgtcgaattgtgtatgagacacacatataaataatgtaacgaagaagtttggagc 686
 QY 826 taccgtcgaagaatttccatcagatgactgtgatacacaatttcccccaaatattgtc 885
 DB 687 tatgttgaacaatcatattatcagatgactgtgatacacaatttcccccaaatattgtc 746
 QY 886 caatcccccgtatttacaacattcacaataatgataatgataacggttattgtt 945
 DB 747 cagttccacgtctgtatgtctgacgtacacttaactaagtgtatgttcaagaagtgctt 806
 QY 946 ggtgatttatacctcctcaagaagtgtaacgaagatcctgttctggaacacacactg 1005
 DB 807 ggtgatttatacctcctcaagaagtgtaacgaagatcctgttctggaacacacactg 866
 QY 1006 aagaagaattccctcggcgagaggtcttcaagatgttctataaagaacccggttccaag 1065
 DB 867 aagaagacagatccctgagggaaggaacgatcatccttcagaagaatattctcccaaga 926
 QY 1066 tctgataattgttatacgaatttctgtgacccatcattgaagtctgtcgaatagtt 1125

RESULT 11
 AAS01169
 ID AAS01169 standard; cDNA; 1642 BP.
 XX
 AC AAS01169;
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Fertilisation-independent endosperm cDNA clone p0104.cabn62r.
 XX
 KW Fertilisation-independent endosperm; plant reproduction; apomixis; ss;
 XX seed; pharmaceutical; nutraceutical; polymer; p0104.cabn62r; corn.
 XX
 OS Zea mays.
 XX
 FT Key Location/Qualifiers
 FT CDS 29..1414
 FT /tag= a
 FT /product= "Fertilisation-independent endosperm protein"

XX WO200116325-A2.
 PN
 PD 08-MAR-2001.
 XX
 PF 30-AUG-2000; 2000MO-US23735.
 XX
 PR 31-AUG-1999; 99US-0151575.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
 PI Simmons CR, Weng Z;
 XX
 DR WPI. 2001-244407/25.
 DR P-PSDB; AA000322.
 XX
 PT New plant fertilization independent endosperm protein for the
 PT production of seed without fertilization is recombinantly produced -
 XX
 PS Claim 7; Page 60; 94pp; English.
 XX
 CC The sequence represents the coding sequence of fertilisation-
 CC independent endosperm clone p0104.cabn62r. Fertilisation-independent
 CC endosperm proteins are plant reproduction proteins necessary for
 CC apomixis, the formation of seeds without fertilisation. Apomixis is
 CC especially useful to agriculture because it eliminates the necessity of
 CC selling plants to produce genetically identical seed. Such seed is useful
 CC to produce seeds for human and animal food and for commercial milling and
 CC extraction, including the production of useful recombinant products in
 CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and
 CC polymers. Embryoless seed production by transgenic plants is less likely
 CC to cause ethical and environmental concern over transgenic plant
 CC production as no gametes are being produced to cross pollinate with other
 CC crops and the seeds cannot germinate.
 CC
 CC Sequence 1642 BP; 421 A; 369 C; 435 G; 417 T; 0 other;

OY	214	ttgaagcaacgacatcaagctctggttgccttgcacccaatgaagaataatcgaaagaatg	273
Db	332	ttaaagatgtagaagtagttagaaggaggttactctgctgcttaagtcggaagcgaagatacaagcc	391
OY	274	acgaaatagatctcagagaaggaagaaaccttctgatacgcttgcttttaactctctgat	333
Db	392	aatagcagaatataactgtgaggaaataacccgatactatgacatcggtgttaacttaattgaac	451
OY	334	gctcgtttcttcgaatgctgcttcgtttacccgctggtgnaaacgagatatactctgtacaaatgct	393
Db	452	atgcgctactatagtctgctttgccttcacgcgcgaattgacatagtgatataattacccgatgc	511
OY	394	ctcgagaatggtggtccataatcagatctgcaatctctatgcctgataagaaataagaagatgc	453
Db	512	cttgagaaatgctggtttcttgctcttcacaanaattatgattgataaggaataagaatga	571
OY	454	ttttacaacggtaaagttaggc---gtgtyggtttaatgggaacccataatgctgcgtga	510
Db	572	ttctaacctctaagctggaaccatcgatcaaatgtatagctaccgcgtctgttgtgctgcga	631
OY	511	ggagtaaaagatacaatcaccgaagtcattgacgtcaacacagtgaaacgattatcaagatctt	570
Db	632	ggaagcaatcgtgatacttcggatccatgacaaattgctgataccgaanaagttagataagctta	691
OY	571	gttggtcatatggaatttaagtgaaacgaataacgaacacaaacctttaaacctaactgtg	630
Db	692	gttggtcattgtgttttaataacatactgataaggaattacacatgcgtctcgaaagccatactac	751
OY	631	atatacgtctagaacaggaatgatactggttcgttggttggaatggttgaaactctgagatgata	690
Db	752	attctctgcagcaaggaatgatactatataagctataaggaatgtaacatccagtcggatctgac	811
OY	691	ttgataattgctgtagagcttggaaggtacacgcataagaaagtctaaagtgtgagattttacacg	750
Db	812	ttagcttttgcaagggctgtgagggcatcgacatgatagtgttgagtggttgaacttccacct	871
OY	751	tcgtatatttaacgcgttctgctagtgtgtgtatgtagacacacataataatgctcaatg	810
Db	872	accgaaggtttggaattttctgcaaatgttggtgcataacatactgtgaanaatttgcataatg	931
OY	811	aaagaagtttggagaatgcgcgcgaagaagctcaacacatgacatgataccatacaaatc	870
Db	932	aaagaattctggataatactgtttgnaaatacctatctatctatgacgtcggaatccatacaaatg	991
OY	871	cccccaaatgttgccaatccctctgatacttatacagctctccatccatacaaatatagat	930
Db	992	ccaaagcggaataatcacgaatcttcgcgtcttgacgtcgacgaagtaaacctgtacatagtgat	1051
OY	931	tgtaacggttggttggttgatcttatactctctcaaaagatgtggaacaagatccctgttg	990
Db	1052	tgtaacaagatgctgtgtgtaacttccatccatacaagaagtgtlaagaagatgcagtttgctt	1111
OY	991	tggaataacacaaactgaaagaagaattctctctgctgagggagcgcttcagaatgttataaga	1055
Db	1112	tggaataacaaacaaacgaagaacgctgtaggcctgtaggaagggagatgtctatgttcttcaaga	1171
OY	1051	taccgggttccaatgtagtatacttgattatcaaatgttctctctgtaacctccatagat	1110
Db	1172	taccgggtgcacaagaatgttcatatagttatgataaatttcaatgctgatttcaactccaac	1231
OY	1111	tcgtgtgcgaabgtaactacgaagaagaaggttatactgtctggaatttgaaagtgtgcct	1170
Db	1232	cagaatggaataagccaacaataaagcgcagatctatgctctggaagtgtgacgtccagccg	1291
OY	1171	cctgtttgtttaaagaagtataacgaatcaatcaaatcaaatctgtaatacggnaaacacgc	1230
Db	1292	cccgctcttaattgacgcgcgtgtcaaacacgaagaatgcaatgcgcgataagagacagccga	1351
OY	1231	atgctgtctgatactgaaacacgaatctctgctgtgtgtagaagggagacatactagctgcgtg	1290
Db	1352	gtgtcatctcgacggaaagcaacgatacccttcttgtagccgcgcgaagaagcgcgcgatactg	1411
OY	1291	gaacg 1294	

```

Db      1412 gacg 1415      |||||
                                |||||
RESULT 13
ID      AAS01165
       AAS01165 standard; cDNA: 1104 BP.
XX
AC      AAS01165;
XX
DT      12-SEP-2001 (first entry)
XX
DE      Fertilisation-independent endosperm cDNA clone p0003.cgpf34rb.
XX
KW      Fertilisation-independent endosperm; plant reproduction; apomixis; ss;
       seed; pharmaceutical; nutraceutical; polymer; p0003.cgpf34rb; corn;
       EST; expressed sequence tag.
XX
OS      Zea mays.
XX
Key      Location/Qualifiers
FH      1..823
FT      CDS
       /*tag= a
       /partial
       /note= "Contains no stop codon"
       /product= "Fertilisation-Independent endosperm protein"
XX
PN      MO200116325-A2.
PD
PD      08-MAR-2001.
XX
PF      30-AUG-2000; 2000WO-US23735.
XX
PR      31-AUG-1999; 99US-0151575.
XX
PA      (DUPO ) DU PONT DE NEMOURS & CO E I.
PA      (PTON-) PIONEER HI-BRED INT INC.
XX
PI      Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
PI      Simmons CR, Weng Z;
XX
PI      WPI: 2001-244407/25.
DR      P-PSDB; AAU00318.
XX
PT      New plant fertilization independent endosperm protein for the
       production of seed without fertilization is recombinantly produced -
XX
PS      Claim 7; Page 53-54; 94pp; English.
XX
CC      The sequence represents the coding sequence of fertilisation-
CC      independent endosperm clone p0003.cgpf34rb. Fertilisation-independent
CC      endosperm proteins are plant reproduction proteins necessary for
CC      apomixis, the formation of seeds without fertilisation. Apomixis is
CC      especially useful to agriculture because it eliminates the necessity of
CC      selfing plants to produce genetically identical seed. Such seed is useful
CC      to produce seeds for human and animal food and for commercial milling and
CC      extraction, including the production of useful recombinant products in
CC      the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and
CC      polymers. Embryoless seed production by transgenic plants is less likely
CC      to cause ethical and environmental concern over transgenic plant
CC      production as no gametes are being produced to cross pollinate with other
       crops and the seeds cannot germinate.
XX
SO      Sequence 1104 BP; 307 A; 204 C; 274 G; 319 T; 0 other;

Query Match      26.0%; Score 406.4; DB 22; Length 1104;
Best Local Similarity 69.9%; Pred. No. 1,1e-109;
Matches 548; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

511 ggcgtaaaggtatataatccgagtcattgacgtcaacagtggaagatcatcataaggtcctt 570
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 ggaacaaatgggatacttcgggtcatcatacttgctgctacagaagaagttgactaaagcttt 70

```

[illegible]

Db 375 ctatgtgatgtacaaagatgctgtgtactcatccatcaagaagtgttacaatga 434
Qy 981 gatcctgtgtgggaacacacactgaagaatctcctgtgcggggagcttgaatgt 1040
Db 435 aatgtgtcttgggaacgaagaacaaagacagatcctgtgggaaagaacatc--gata 492
Qy 1041 tctatgaatataccggttccaatgtgtatattgtttatcaagtttctgtgaact 1100
Db 493 tcttcagaagaactgtctccagaatgtgacatttgggttatcaaatcttcatgtgatt 552
Qy 1101 ccatttaagtctgtgtcgatagtgtaatcagaagga 1139
Db 553 tcaattca-atcagtcgcgataggaaccgtaagaacaa 590

RESULT 15
AAS01189
ID AAS01189 standard; cDNA; 534 BP.
AC AAS01189;
AT AAS01189;
CT 12-SEP-2001 (first entry)
DE Fertilisation-independent endosperm cDNA clone ses2w.pk0015.b10.
XX Fertilisation-independent endosperm; plant reproduction; apomixis; ss;
KW Seed; pharmaceutical; nutraceutical; polymer; soybean; ses2w.pk0015.b10;
KM EST; expressed sequence tag.
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 2..489
FT /tag= a
FT /partial
FT /product= "Fertilisation-independent endosperm protein"
FT /note= "Contains no start or stop codon."
XX
PN WO200116325-A2.
XX
PD 08-MAR-2001.
XX
PE 30-AUG-2000; 2000MO-US23735.
XX
PR 31-AUG-1999; 99US-0151575.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
PI Simmons CR, Weng Z;
XX
XX WPI; 2001-244407/25.
DR P-PSDB; AAU00342.
XX
XX
PT New plant fertilization independent endosperm protein for the
PT production of seed without fertilization is recombinantly produced -
XX
XX
PS Claim 7; Page 88-89; 94pp; English.
XX
XX
XX The sequence represents the coding sequence of fertilisation-
XX independent endosperm clone ses2w.pk0015.b10. Fertilisation-independent
XX endosperm proteins are plant reproduction proteins necessary for
XX apomixis, the formation of seeds without fertilization. Apomixis is
XX especially useful to agriculture because it eliminates the necessity of
XX selfing plants to produce genetically identical seed. Such seed is useful
XX to produce seeds for human and animal food and for commercial milling and
XX extraction, including the production of useful recombinant products in
XX the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and
XX polymers. Embryoless seed production by transgenic plants is less likely
XX to cause ethical and environmental concern over transgenic plant
XX production as no gametes are being produced to cross pollinate with other
XX crops and the seeds cannot germinate.

XX
SQ Sequence 534 BP; 155 A; 86 C; 118 G; 175 T; 0 other;
Query Match 19.1%; Score 298.6; DB 22: Length 534;
Best Local Similarity 74.5%; Pred. No. 5.7e-78;
Matches 376; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
Qy 561 taagatcttgggtgacatgagatcgaatgaagaacatcagacacacatttaaac 620
Db 1 taagatcttgggtgacatgagatcgaatgaagaacatcagacacacatttaaac 60
Qy 621 tcaactgtgatactcgtcgaagaatgaatcgttctgttgggaagtgaactcgtg 680
Db 61 atcactcgtgatactcgtcgaagaatgaatcgttctgttgggaagtgaactcgtg 120
Qy 681 gataatgatttgaatattgtcgtgagcgtgaggtacacgtcgaagtcttaagtga 740
Db 121 aatatgatttgaatattgtcgtgagcgtgaggtacacgtcgaagtcttaagtga 180
Qy 741 ttctacccgtctgataatttaccgcttctgtagttgtgtatggacacacatttaat 800
Db 181 ttctacccgtctgataatttaccgcttctgtagttgtgtatggacacacatttaat 240
Qy 801 atgtcaatgaagaagtgttggacgtacgttcgagaagtcattcacaatgacatgacc 860
Db 241 atgtccttaagaagaagtgttggacgtacgttcgagaagtcattcacaatgacatgacc 300
Qy 861 atcaaatcccccacaaaatttggccaattccctgtatttaccagcttccattacaacaa 920
Db 301 ttccaagtttccacaacaaatttggccaattccctgtatttaccagcttccattacaacaa 360
Qy 921 ttatgtagattgaaccgttgggttggatattatccctccaagaagtgtgacaacga 980
Db 361 ttatgttagctgtgaataggttgggttggatattatccctccaagaagtgtgataatga 420
Qy 981 gatccgttctgggaaccccaactgaaagaattctcctgtggaaggaggttccaatgt 1040
Db 421 aatatccttgggtgaaccctcaaatgtaacgaaccaactccaagggaaggtgttagtgtat 480
Qy 1041 tctatgaatataccggttccaatg 1065
Db 481 cctcataaatacccatcttcgattg 505

Search completed: June 25, 2002, 00:17:22
Job time: 5356 sec

Description

[illegible]

RESULT 1
AR150926

DEFINITION Sequence 3 from patent US 6229064.

VERSION ARI50926.1 GI:15115517

SOURCE Unknown.

Unclassified.

AUTHORS Fischer, R.L., Ohad, N., Kiyosue, T., Yadegari, R., Margossian, L.,

TITLE Nucleic acids that control endosperm development in plants

FEATURES

Location/Qualifiers

/organism="unknown"

ORIGIN

Query Match	100.0%;	Score 1563;	DB 6;	Length 1563;
-------------	---------	-------------	-------	--------------

Db	Accession	Gene	Length	Source
Db	1021	GGCGAGGAGCTTCACATGTTCTATTAAAGAACCCGGTCCAAATGTGAGATATTGGTTT	1080	
QY	1081	atcaagttcttccttgagccctccalttaagtcctgtgcgataaggtaatcatcagaagaag	1140	
Db	1081	ATCAAGTTTCTTGAGACCTCCATTAAAGTCTGTGGATGAGTAAATCAGAAGGAAG	1140	
QY	1141	gtttatgtctcggatggaagaaagttgccccttcgttttgattacaagaagtaacacaat	1200	
Db	1141	GTTTATGTCTCGGATGTTGAAAAGTTGGCCCTCTGTTTGATTACAAATTATCACACAT	1200	
QY	1201	caatcaagctctgtaatatcaagcaacacagcaatgctcgtcgataggaaagacattctgct	1260	
Db	1201	CAATCAAGTCTGTRATATAGCCAAACACCANTGTCTGTGATGGAAGCAGCATCTTGTCT	1260	
QY	1261	tgctcgagagacgagactatattgagcgttgaggacgtgattaccgaagtagcggtcgaagct	1320	
Db	1261	TGCTCGAGGAGCAGGACTATATGCGGCTGGGACGTGATACCAAGTACCGGTCTGAGTCT	1320	
QY	1321	tgtagaatctgaatgaattagaagatgctgagaagaatgaagtaaccattctttattgtaatt	1380	
Db	1321	TGTAGAAATGTTGTAATTAAGATGAGTCCGAAGAAATGATATCCATCTTTTATTTGTAATT	1380	
QY	1381	ctgatacattgctactctccctgagacccttgagatgctctttgtaagccttgtaacgtcca	1440	
Db	1381	CTGATCATGTTGTTCTACTCCCTGAGACCTTGAGATCTCTTTGTACCCTTGTTAAAGTCCA	1440	
QY	1441	cccttgataccacagtgatatacccttccttgagagatttgctcttattctcttagttcaat	1500	
Db	1441	CCCTTGATCACACAGTGTATACCTTTCTGAGAAATTTGCTTATTTCTCTTTAATTTCAATAC	1500	
QY	1501	acaagctgataccctggagccttatttgaggagaaccactctcttataagcttcctagta	1560	
Db	1501	ACAAGCTGTATCTGAGACTTATTGACAGAACACACTCTCTTTATTAAGCTTTCTAGTA	1560	
QY	1561	ttt	1563	
Db	1561	TTC	1563	
RESULT 2				
AF129516		1563 bp	mRNA	linear
LOCUS				PLN 06-APR-1999
DEFINITION				Arabidopsis thaliana fertilization-independent endosperm protein
ACCESSION				AF129516
VERSION				AF129516.1
KEYWORDS				GI:4567094
SOURCE				
ORGANISM				thale cress.
				Arabidopsis thaliana
				Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
				Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
				Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE				1 (bases 1 to 1563)
AUTHORS				Ohad,N., Yadgarli,R., Margossian,L., Hannon,M., Michaeli,D.,
				Harada,J.J., Goldberg,R.B. and Fischer,R.L.
TITLE				Mutations in FIE, a WD polycarb group gene, allow endosperm
				development without fertilization
JOURNAL				Plant Cell 11 (3), 407-416 (1999)
MEDLINE				9917205
REFERENCE				2 (bases 1 to 1563)
AUTHORS				Ohad,N., Yadgarli,R., Margossian,L., Hannon,M., Michaeli,D.,
				Harada,J.J., Goldberg,R.B. and Fischer,R.L.
TITLE				Direct Submision
JOURNAL				Submitted (18-FEB-1999) Plant and Microbial Biology, University of
				California, 111 Koshland Hall, Berkeley, CA 94720-3102, USA
FEATURES				Location/Qualifiers
source				1..1563
				/organism="Arabidopsis thaliana"
				/cultivar="Landsberg erecta"
				/db_xref="taxon:3702"
				/chromosome="3"
				/map="between EMB29 and AXR2"
gene				1..1563

900 AACGAGATTGCTCTTGGGAACCAAAATGAGGAACATCTCCGGGAGAGGGATCGGTG 95

Qy 499 gttcgcgcctgagagatataatccgattgtacgtcaacagtgaacgatt 558

[illegible][illegible]

RESULT	6
AX090444	
LOCUS	AX090444
DEFINITION	Sequence 33 from Patent WO0116325.
ACCESSION	AX090444
VERSION	AX090444.1 GI:13444284
KEYWORDS	
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa
	[MIM:271800] Strachan; Embryo; Tracheophyta;

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1498)
Butler, K.H., Danilevskaya, O., Miao, G.H., Morgante, M., Sakai, H.,
Simmons, C.R. and Weng, Z.
Plant reproduction proteins
Patent: WO 0116325-A 33 08-MAR-2001;
F I DIT POINT DE MEMOIRS AND COMPANY (US) ; Pioneer Hi-Bred

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Asterophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

FEATURES	source	unsure	BASE COUNT	ORIGIN
International, Inc. (US)				
Location/Qualifiers				
1. .1498				
/organism="Oryza sativa"				
/db_xref="taxon:4530"				
839				
411 a	304 c	356 g	426 t	1 others

Query Match	33.7%;	Score 526;	DB 6;	Length 1488;
Best Local Similarity	68.6%;	Pred. No. 1,2e-127;		
Matches 724;	Conservative	0;	Mismatches 331;	Indels 0;
			Gaps	0

Oy	240	tttgaatccatcgcataagaatcgtacaagaagtacgaataggtttccaaagaggaaaga	299
Db	53	ttttggtttccaaagcagggaagggagttcaaaaggcgtccaaacaaagctccacaggggacgc	112
Oy	300	accttctgatgctgtgttttccaacttccttgatgcctgttcttgatgcttcctgctac	359
Db	113	gcacctctaaccccttcggattccaattcttcgcattccactactacgaggcttttcgccac	172
Oy	360	cgctcggtggaaatcggtactcgtgaacatgtctcgagatggtgcataatcaat	419
Db	173	cgctggccgcacacgcgcggacacacctcaactcagctccctcaagatggtatttttgctatnct	232
Oy	420	gcaatcccatgctgtagaagaataaggaagaagtcgttttacaacgttaagtgtggcgtgtg	479
Db	223	gcaaccatattattgtatgatgaattaggaatgcgtttctacacactgaatttggcctttgtga	292
Oy	480	cgttaatvggaaaccatattgttcggctcggaggagtaaaaggttaatccgaagtcattga	539
Db	293	tctttgatggcacacccgctgttagtggctgcacaggaagcaatgggaatcattccgggtatcaa	352
Oy	540	cgtcaacagtgtaaacagattcataaagatcgttgggtcatgtgagattcaagtgaacgaat	599

Db 653 CTTTCATGAGCATGACCTTCATCCAAATTTCCAAACAAATAATATGCAATTTCCGACCTA 772

Qy 900 taccagctccatccattacaaattatagatcttgaacccgttggtttggtatattaccc 959

Db 713 GGTTGCTGAGACACTTCTACTACTAGTTGATTTGATGACGTGAGCGCTTGATGACTTCATTCT 772

Qy 960 ctcaagaagtgtagcaacagagatccctgttcttggaacccaacttggaagaaattcc 1019

Db 773 GTCACAGAGATCTTGACATGAATTTGCTCTGGGAGGCCCAAAACAAAGAACAAAGTCC 832

[illegible][illegible]

Qy 1260 ttgctgcaggaacggagactatcggcgccgggacg 1274
||| ||||| | | ||||| |||||
Db 1073 ctgcaggcagagatgccacgatattggcgatggcatg 1107

RESULT	7				
AX090448					
LOCUS	AX090448	1518 bp	DNA	linear	PAT 21-MAR-2001
DEFINITION	Sequence 37 from Patent WO0116525.				
ACCESSION	AX090448				
VERSION	AX090448.1	GI:13444286			
KEYWORDS					
SOURCE	bread wheat.				
ORGANISM	Triticum aestivum				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
	Poideae; Triticaceae; Triticum.				
	1 (bases 1 to 1518)				
	Butler,K.H., Danilevskaia,O., Miao,G.H., Morgante,M., Sakai,H.,				
	Simmons,C.R. and Weng,Z.				
	Plant reproduction proteins				
	Patent: WO 0116525-A 37 08-MAR-2001;				
	E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred				
	International, Inc. (US)				
	location/Qualifiers				
	1..1518				
FEATURES					
SOURCE					

/organism="Triticum aestivum"
BASE COUNT 393 a 337 c 403 g 384 t 1 others
ORIGIN

Query Match 31.9% Score 498.2; DB 6; Length 1518;
Best Local Similarity 65.3%; Pred. No. 2.5e-120;
Matches 731; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

214 tttaggaacgaatcaatgaatgttggtcttctgacatcgaataagaacgtacaaagtcg 273
115 tttaggggtggagagcgccggctgggtgcctggcccccacccggagccggagatcaacgctc 174
274 acgaatagatcaggaaggaagaacacctgtatgtctgtgttttcaactccttgat 333
175 ttcacacgaacacacgagggcgaagccgcctctctagcccatggcttcaacttctacgac 234
334 gctgttctctgagatgtctgttaccgtgtgtgaatacsgatctctgttacaattgtc 393
235 gcccccttactagctgttccgcacccgctggcgaatcgtgtgacgacgacgacgacg 294
394 ctccgagatgtgtccatcacaatcgaatcctatctgtctgtatgaagataagaagtcg 453
295 ctccccgacggaatctgtgctgtctgcaagacatattgatcgagacatcctcactga 354
454 tttaacacggaatgtggcgctgtgcgttaataatggaaacacatagtgcgctgtgagga 513
355 ttctacacactagctgtggctgtgtgacgttgacgacgacacacgctctagtcgacgacga 414
514 gtaaaagatataatccgaatcgtacgtcaacgaagtaagaatcattacgaatgtctgtg 573
415 accaattgccgctatttgcggtctatcactgtccacccgagaagtggttaagagtttctt 474
574 ggtcataggaatcagtgagacgaatacgaacacacacacacacacacacacacacacac 633
475 gcccatgtggtatcattatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 534
634 actgttagcagaatgaatgaatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 693
535 tctgcagacgaagacgagctgtgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 594
694 atattgttcgagctgtgagctgt 753
595 atttttgcctggagagagagctgt 654
754 gatattacacgctgt 813
655 gatattcctacgagatgt 714
814 gagttttagcagtgatcagtgagacacacacacacacacacacacacacacacacacacac 873
715 gaatttttgcctggagagagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 774
874 acaaaatgttccac 933
775 accgaattttgttcaatttttccctctcactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 834
934 aacgctgt 993
835 actagtggt 894
994 gaac 1053
895 gagcccaaaaattaaaagagagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 954
1054 ccggttccaaatgt 1113
955 cctgtgtccctgattgt 1014
1114 gttgtgataagtaacgaaggaaggaaggttattgtctgtgagatttgaagaagttgcctctc 1173
1015 tttagcaatgtgcaacacgacgaagcaaaaattctatgtgtgtgtgtgtgtgtgtgtgtgtgt 1074

Qy 1174 gttttattacaagaattatcacacaatcaaatcaaatgtctgttaatcaggaacagcagtcg 1233
Db 1075 gtgcataatttaccgctgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1134
Qy 1234 tctgtcagtgaaagacagatcttctgtctgtcaggaagcaggaacataatgagcgtggagc 1293
Db 1135 tctgttattgaaagacagatcttctgtctgtcaggaagcaggaacataatgagcgtggagc 1194
Qy 1294 gtgattaccgaatgagcgt 1332
Db 1195 gaagtgaaacatcagctgt 1233

RESULT 8
AX090412
LOCUS AX090412 1643 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 1 from Patent WO0116325.
ACCESSION AX090412
VERSION AX090412.1 GI:13444268
KEYWORDS
SOURCE
ORGANISM
Zea mays.
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 1643)
BUTLER, K.H., DANILEVSKAYA, O., MIAO, G.H., MORGANTE, M., SAKAI, H.,
SIMMONS, C.R. and WENG, Z.
Plant reproduction proteins
Patent: WO 0116325-A 1 08-MAR-2001;
E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred
International, Inc. (US)
Location/Qualifiers
1..1643
/organism="Zea mays"
/db_xref="taxon:4577"

BASE COUNT 429 a 345 c 456 g 413 t
ORIGIN

Query Match 31.3% Score 489.4; DB 6; Length 1643;
Best Local Similarity 66.4%; Pred. No. 5.2e-118;
Matches 719; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

Qy 215 taaggaaacgaatcagtaagttgggtcttctgacatcgaataagaacacgtacaaagtcg 274
Db 221 tgcggtgcagagggcgagagggctgcctgcgcacgacgggaagcggaatgacacccgt 280
Qy 275 cgaatagatcagaaggaaggaacaccttgatgacgtgtgtgtgtgtgtgtgtgtgtgtgtgt 334
Db 281 gcggcgaagacacactgtgaggggaacggcctgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 340
Qy 335 ctgcgttctcagatgtctgtctgtacgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 394
Db 341 cgcgctactacacacgcttctgcacccgctgcgcgacacccgctgtacacacacacacacacac 400
Qy 395 tcggaagtggtgcacatcagaatccatccatctgtatgtatgaagaagaagtcgt 454
Db 401 tttagaattgt 460
Qy 455 tttaacagtgatgt 511
Db 461 tctatgtctgt 520
Qy 512 gagtaaaagatataatccgaatcagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 571
Db 521 gaagcaattggatcatttccgggtcattcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 580
Qy 572 tgggtcagtgaaatcagtgaaacgaatcaggaacacacacacacacacacacacacacacacac 631
Db 581 ttggcctatggcgcactcaataatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 640

Query Match	31.3%;	Score 489.4;	DB 6;	Length 1686;
Best Local Similarity	66.4%;	Pred. No. 5,2e-118;		
Matches 719;	Conservative 0;	Mismatches 361;	Indels 3;	Gaps 1;
Db	215	taggaagacagatcacaatgtgtgtcttcttgactcacaatgaataagaatcgtacaaagtga	274	
Db	422	TCGGGTGGAGGGGGGGGGGGGGTGGCTGCTGCTGCCAGCGGAGCGGGAGTGAAGCCCT	481	
Qy	275	cgaataggatcagaagaaggaaaccttctgtatgcctgtgtttcaactccctatg	334	
Db	482	GCGGCACACACTGATGGGGGAGCCCGCTATATGCTATTCGGGTTCACCTTCANTGAGC	541	
Qy	335	ctcgcttcttcgtatgtcttcgttaccgcctgctggtgaatcggatcttactcgtacattgtc	394	
Db	542	GCGGCTACTACGACGCTCTTCGCCACCGTCGGGGGCAACGGCGGTGAACATTACCGGTGCC	601	
Qy	395	tcggaagatgltgcacatcagatctgtcaatctcctaigtctgaatgaagaagaagtcgt	454	
Db	602	TTGAGAAATGCTAGTTTCGCTCTTTCACAGACTTTACGTTGATGAGTATGATGATGCTGT	661	
Qy	455	tttacacggttaagttgggcgtgtgac---gttaatgggaaccacatagttgtggcctggag	511	
Db	662	TCTATATCTTAAGCTGGGCTCGTGACCAATGTTGATGCGTCCACCACTGCTGGTGACGACG	721	
Qy	512	gaagtaaaagtataatccgaatcatttgaacgtcaacagtgaaacagatctaaagatcctg	571	
Db	722	GAAGCAATGGGATCATTCGGGTCTATCAATTTGCTTACAGAAAAGTTAGCTAAGACGTTTG	781	
Qy	572	tgggttcaggaatctcagcgaacgaatcaggaacacacactttaaactcaactctgtga	631	
Db	782	TTGCCATGGCCAGCTCAATTAATGATGAATAAGAACTCAACGTTGAAGCCCTGCCTCATCA	841	
Qy	632	ttactgtcagcaaggaatgaatctgttcgttctgttggaaatgttgaactgggatgtatct	691	
Db	842	TTTTCTGCACAGCAAGATGATGATCTGTAGGCTATGGAATGCTCCATACAGGGATCTGATCT	901	
Qy	692	tgaatcttctgtgagctgtgaggtcactgcctatgaagtctctaagtcgtgaatttcacocgt	751	
Db	902	TGATATTTGCTGGAGCTGGAGGTCATCGCATGCAATGAAGATTGAATGTTGACTTCCATCCTA	961	
Qy	752	ctgaatcttaccgcctctgtcagttctgtgtatgttgaaacacacatctaaataatgttcaatga	811	
Db	962	GTCATATGAACTTTTTCGCAAGTTGGGCAATGGACATGACAACTGTGAATAATCGTGCATATGA	1021	
Qy	812	aagagttcttgacgtacgtcgcagaaagtcacatcacaatgagctgatgatccatcaaatcc	871	
Db	1022	AAGAAATTTTGGTATATGTTGACAAATCATATTTCAATGATGACCTTCCTCATCAAAAGTTTC	1081	
Qy	872	ccacaaatctgtccaaatccctgtatcttcagcttcacatccatcaataatgtatgact	931	
Db	1082	CAACAAATATGTCACATTTTCAGTTCAGTTCGTCGATGATACACTTAACTATGTTATAT	1141	
Qy	932	gtaacgctgtgtgtgtgtatcttaccctccaaagctgtggaacacagagatccctgttgt	991	
Db	1142	GTCACATATGCTTTGGTGACTTCATCCTATCAAAAGATGTTGACAAATGATGCTT	1201	
Qy	992	gggaacacacaactgaaagaatctctccctgycgagggagcttcagatgctcatlaaagt	1051	
Db	1202	GGGAGCGGAAGACAAAGAACAGAGTCTGGGGGAGGAGAACATGATATCTTCAGAGT	1261	
Qy	1052	accgcgttccaatgtgtgatatttgtttatcaagtttcttctgtgactccatttaagtc	1111	
Db	1262	ATCCTGTCCCAAGATGCAATGCTTTTATCAAAATTTTCAATGATTTTCATGTTCTCATATC	1321	
Qy	1112	ctgtgtgatatagtgaaatcaggaagaagtttactgtcgtgatttggaaagtctgcctc	1171	
Db	1322	AGTTGGGATAGGCAACCGTGAAGGCAAAATCTACGTGTGGGAAGTACAGTCCAGCCCTC	1381	
Qy	1172	ctgttctgatatcaaaagtatcacacaatcatcaatgaagctgttaatcaggaacaaagcca	1231	

QY	812	aagagcttctggacgctacgctcgaagaagctatcacaatgactgacatgacatccatcaaatcc	871
Db	1118	AAGAAATTTGGCTATATGTTGACAAATATATCATGTGACTGACACTTCATCAAGATTTTC	1177
QY	872	ccacaaatctgtccaatctccgttatttacaagttcatcattcacaatatagtatgt	931
Db	1178	CACAAAAATATGTCACAGTTTCCAGTCTTGATTTGCTGGCAGTACACTTACACTATTTGATTT	1237
QY	932	gtaaccttggcttggatgtatttaccctcacaagaagtgtgacacacagatccctgtgt	991
Db	1238	GTAACAAGATGGCTTGTCGTCACATCTTGATTTGCTGGCAGTACACTTACACTATTTGATTT	1297
QY	992	gggaaccacacactgaagaagaattccctcggcggaggagcttcagatgttctatlaagat	1051
Db	1298	GGGAACCGAAGACAAAGAACAGAGTCTCGGGGAGGGAACGATCATCTTCAGAGAT	1357
QY	1052	acccggttccaatgtgtgatatattggtttaccagaattctctgtgacctccattaaagt	1111
Db	1358	ATCCGTCGCCAGAAATGTACATTTGGTTTATCAAAATTTTCAATGTGATTTTACCTTCAATTC	1417
QY	1112	ctgttgccgatggtaaatcaaggaagaagttatgtctcggatctggaattggaattgcctc	1171
Db	1418	AGTTGGCCATATGCGCACCGCTGAGGCAAAATCTACGTGTGGGAAGTACAGTCACGCCCTC	1477
QY	1172	ctgttttattacaagaattacacacaatcaatcaaatcaagctctgatatcagcgaacagcca	1231
Db	1478	CTGTCTCATTTGCTCGGCTGTATATATCAGCAGTGTAAATCCCGCATAGACAAACTCGAG	1537
QY	1232	tgctctgcgatggaagcagcagattctgtctcgtcggagaagcaggactataggcgctggg	1291
Db	1538	TGTCCTTCGATGGAGGACCAATCTTGGAGCTGTGTGAAGCGGACCATCTGGCGGTGGG	1597
QY	1292	acg 1294	
Db	1598	ATG 1600	
LOCUS	AX090428	1486 bp	DNA
DEFINITION	Sequence 17 from Patent WO0116325.		linear
ACCESSION	AX090428		
VERSION	AX090428.1	GI:13444276	
KEYWORDS			
SOURCE			
ORGANISM	Zea mays.		
REFERENCE			
AUTHORS	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.		
TITLE	1 (bases 1 to 1486)		
JOURNAL	Butler,K.H., Danilevskaya,O., Miao,G.H., Morgante,M., Sakai,H., Simmons,C.R. and Weng,Z.		
FEATURES	Plant reproduction proteins		
SOURCE	Patent: WO 0116325 A 17 08-MAR-2001;		
BASE COUNT	E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred International, Inc. (US)		
ORIGIN	Location/Qualifiers		
	1..1486		
	/organism="Zea mays"		
	/db_xref="taxon:4577"		
	391 a 306 c 380 g 408 t 1 others		
Query Match	30.2%;	Score 472.2;	DB 6; Length 1486;
Best Local Similarity	68.1%;	Pred. No. 1.7e-113;	
Matches 687; Conservative	0;	Mismatches 318;	Indels 4; Gaps 2
QY	289	gaagggaagaacacttgatctgtctgttcttccaactcttgatgctgttcttctgat	348
Db	147	GAGGGAGACGCCCGCTATATGCTATGCGGTTCAACTTATGTAGAGCGCGGCTACTACGAC	206

OY	454	ttttaaccggaagcttgggc---	gtgtggcgcttaatgaggaaacccataatgttcgctgtga	510
Db	572	TTCTACACTC	TAAAGCTGACCATGCATCAAGTGAATGAGCTCACCGGTGTGGTGCCGCA	631
OY	511	ggagttaaaggatataatccgaatcattgaagtgcaacg	tgaaacgattcaagaagtcct	570
Db	632	GGAAGCATTCGGATCATTCGGGTATCAATTGCTACCCGAAAGTTAGATTAAGAGCTTA	691	
OY	571	gtgggtcatgtagatatgaagaaacaatccagaaacaaccttaaaacctcaactgtg	630	
Db	692	GTTGGCCATGTGGTGTTCANATACATGAGATAAGACATCATGCTCGTAGAAGCATCAGTCATC	751	
OY	631	attaactgtagaagaagatgaatcctgttcgttttgtgaaatgttgaactvgatgat	690	
Db	752	ATTTCCTGCCACGAAGATGATGATATATTAGCGTATGGAATGCATCTACTGTGGATTTGGATC	811	
OY	691	ttgatatattcgtgtgaagctggaagtcataatgcatgaagttctaagtgtagatttcacgcg	750	
Db	812	TTAGCTTTTGGCAGGGGCTGGAGGCGCATCGACATGATGTTGAGATGTGGACTTCACCTT	871	
OY	751	tctgattattacccgtttgtcgtatgtgtgtaattggaaccaactataaataatgltcaatg	810	
Db	872	AACCGAGTGTGGGATTTTTTGGCAAGTGTGTGGCATTTGGCANATCTGTGAAAATTTGGTGCATG	931	
OY	811	aagaagatttgagacgtacgtcgaagaagcatccaatgagatgagatgatalccataaatlc	870	
Db	932	AAGAATTTTGGATATATGTTGAAAAATCATATTCATGATGAGCCATCCATCAAAGTTT	991	
OY	871	cccacaaaattgtccaattccctgttatattacagcttccatccaataaataatgagat	930	
Db	992	CCAAGGAGAAATATCCAGTTTCCGGCTTGGACTGTCAGTACACTGTGACTATGTTGAT	1051	
OY	931	tgtaacctgtgtgtgtgtgtatattacccccaagaagtgtggaacaagagatcctgtg	990	
Db	1052	TTTACAGATGTGCTTGTGTGACTTCACTCTATCAAGAAGATGAAGAATGACGATTTGGCTT	1111	
OY	991	tgggaaccacaactgaagaagaattctcctgtggaaggagcttcagatgltctaltaaga	1050	
Db	1112	TGGGAACCAAAACAGACAAGCGTGAAGCGCTGGGGAGGGAGTGTGATGTTCTTCAGAAAG	1171	
OY	1051	taccgggttccaaatgtgtgataattgtgtttalcaagtttctgtgacctcatltaagt	1110	
Db	1172	TACCCGGTCCCAAGAGTTCATTATGAGTTATGAATAATTTCAATGTGATTTTACTCCAAC	1231	
OY	1111	tctgtgtgcatgtgtaataagaagaagaagtgtatcgtgtgagattgaaaaagttgccct	1170	
Db	1232	CAGATGGCATATAGGCAACAATAAAGCGCAGATCTATGCTGTGGAAAGTGCAGTCCAGCCGG	1291	
OY	1171	cctgttttgattacaagatcataccaataccaatcaaatcgtatcaatcaaggcaaacagcc	1230	
Db	1292	CCCCTCTTAATTGACCGGCTGTGCCAACCAAGAAATGAATGCCCATTAAGCAGACCCGA	1351	
OY	1231	atgtcgtcgtgtagaagacgaatcttctgtcgtcgtcgagagcggagataatgagcgtcgg	1290	
Db	1352	GTTGTATTCGACGAGCAGACGATCTTGGAGCGCCGACGACGGCGCATGTCGGGGGTGG	1411	
OY	1291	gacg 1294		
Db	1412	GACG 1415		
RESULT	15			
LOCUS	AY061964	1787 bp	mRNA	linear PLN_01-JAN-2002
DEFINITION	Zea mays fertilization-independent endosperm protein (FIEP)		mRNA,	
ACCESSION	AY061964		complete cds.	
VERSION	AY061964.1	GI:18032003		
KEYWORDS	.			
SOURCE	Zea mays.			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			


```

|||||
Db 181 agagattcgaatcgaatgctgaagaataccttagggaacgaagtcgaatgctggtct 240
Oy 241 ttgaccctcgataagaatcgtacaagaatgacgaataggaattcaagaaggaga 300
Db 241 ttgaccctcgataagaatcgtacaagaatgacgaataggaattcaagaaggaga 300
Oy 301 ccttctatgctgtcttctcaactcctctgacgtcgtcttctctgacgtctctacc 360
Db 301 ccttctatgctgtcttctcaactcctctgacgtcgtcttctctgacgtctctacc 360
Oy 361 gctggtggaatcgtgactctgtacaaatgtctcgagaatggtccataacgacttg 420
Db 361 gctggtggaatcgtgactctgtacaaatgtctcgagaatggtccataacgacttg 420
Oy 421 caatcctatgctgataagaatggaagagctcttcaacggttaattggtggtggtg 480
Db 421 caatcctatgctgataagaatggaagagctcttcaacggttaattggtggtggtg 480
Oy 481 gttcaatggaaccataatgtctggtcggtggaagtaaaagtatacgcagtcattgac 540
Db 481 gttcaatggaaccataatgtctggtcggtggaagtaaaagtatacgcagtcattgac 540
Oy 541 gtcaacagtgaaaacgatcacaagagtcctgtggtggtcgtacgtgacgtacgaatc 600
Db 541 gtcaacagtgaaaacgatcacaagagtcctgtggtggtcgtacgtgacgtacgaatc 600
Oy 601 aggaacaaaccttaaaactcaactgtgattactctgctggaagtgatcgtctgct 660
Db 601 aggaacaaaccttaaaactcaactgtgattactctgctggaagtgatcgtctgct 660
Oy 661 ttgtggaatgtgaactggaatgataatgatttgaatttctgagtcgtgagtcacgc 720
Db 661 ttgtggaatgtgaactggaatgataatgatttgaatttctgagtcgtgagtcacgc 720
Oy 721 tatgaatcttaagtggtgatttcaatcgtcgtatattacgcgttctgactgtggt 780
Db 721 tatgaatcttaagtggtgatttcaatcgtcgtatattacgcgttctgactgtggt 780
Oy 781 atggaacccactataaataatgtaataagaagtttggaagtcagtcgaagaatga 840
Db 781 atggaacccactataaataatgtaataagaagtttggaagtcagtcgaagaatga 840
Oy 841 ttccatgagactgtatccatccaaatccccaacaaatttgcacaaatccctgatt 900
Db 841 ttccatgagactgtatccatccaaatccccaacaaatttgcacaaatccctgatt 900
Oy 901 acagctccatcacaataatgtagatgtgaacgtgtgtgtgtgtgtgtgtgtgtgt 960
Db 901 acagctccatcacaataatgtagatgtgaacgtgtgtgtgtgtgtgtgtgtgtgt 960
Oy 961 tcaagaagtggtggaacagagatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
Db 961 tcaagaagtggtggaacagagatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
Oy 1021 ggcgaaggagcttcagatgtctataaagataccggttccaaatgtgtgataattg 1080
Db 1021 ggcgaaggagcttcagatgtctataaagataccggttccaaatgtgtgataattg 1080
Oy 1081 atcaagtttctgtgacccatcaatgaatgtctgtgtgtgtgtgtgtgtgtgtgt 1140
Db 1081 atcaagtttctgtgacccatcaatgaatgtctgtgtgtgtgtgtgtgtgtgtgt 1140
Oy 1141 gtttatgtcgtgatttgaagaatgtgctcctcctgttctgtatatacaaatgtac 1200
Db 1141 gtttatgtcgtgatttgaagaatgtgctcctcctgttctgtatatacaaatgtac 1200
Oy 1201 caatcaaaagctgtatcaagaacaaacagcagatgtctgtgtgtgtgtgtgtgtgt 1260
Db 1201 caatcaaaagctgtatcaagaacaaacagcagatgtctgtgtgtgtgtgtgtgtgt 1260
Oy 1261 tgcgtggaagcaggaactatagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320

```

```

Db 1261 tgcgtggaagcaggaactatagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
Oy 1321 tgtagaatgtatgaatgtgagtcggaagaatgataatccattcttattgttaatt 1380
Db 1321 tgtagaatgtatgaatgtgagtcggaagaatgataatccattcttattgttaatt 1380
Oy 1381 ctgatatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
Db 1381 ctgatatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
Oy 1441 cccctgtacacagtgatatacccttctgtgagatttgttcttattcttattgtta 1500
Db 1441 cccctgtacacagtgatatacccttctgtgagatttgttcttattcttattgtta 1500
Oy 1501 acaagctgtatcctgtgagcttattatggaagaacacactcttcttaagcttct 1560
Db 1501 acaagctgtatcctgtgagcttattatggaagaacacactcttcttaagcttct 1560
Oy 1561 ttc 1563
Db 1561 ttc 1563

```

RESULT 2

```

US-09-177-249-5
; Sequence 5, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramon
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; FILE REFERENCE: Development in Plants
; CURRENT APPLICATION NUMBER: US/09/177, 249
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5801
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3872)..(5566)
; OTHER INFORMATION: Fertilization-independent endosperm 3 (FIE3)
; OTHER INFORMATION: WD40/polycomb gene genomic sequence
US-09-177-249-5

```

```

Query Match 24.0%; Score 375.8; DB 4; Length 5801;
Best Local Similarity 99.5%; Pred. No. 8.5e-105;
Matches 377; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 aaagtgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 60
Db 2319 aaagtgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2378
Oy 61 acggttacgaatattataaaccgtgatactgtgtacatcgaaatatacaaacacc 120
Db 2379 acggttacgaatattataaaccgtgatactgtgtacatcgaaatatacaaacacc 2438
Oy 121 agtgtgtgtcatgagaactgtgagcttagatagatgtgtgtgtgtgtgtgtgtgtgtgt 180
Db 2439 agtgtgtgtcatgagaactgtgagcttagatagatgtgtgtgtgtgtgtgtgtgtgtgt 2498

```


QY	181	agaatttcgaatcgaatcgaatgctcgaatgaacgaatgaatggtggtc	240
Db	2499	aggaatttcgaatcgaatcgaatgctcgaatgaacgaatgaatggtggtc	2558
QY	241	ttgaattccatcgaataagaatcgtacaagaatgacgaatgattcagaagaaggagaa	300
Db	2559	ttgaattccatcgaataagaatcgtacaagaatgacgaatgattcagaagaaggagaa	2618
QY	301	ccattgtaatgctgtgtttccaactccctgatgctcgattcctcgatgctctgtacc	360
Db	2619	ccattgtaatgctgtgtttccaactccctgatgctcgattcctcgatgctctgtacc	2678
QY	361	gctgtgtaagaatcgattaa	379
Db	2679	gctgtgtaagaatcgattaa	2697

```

US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHIEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOMLOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 diagonal Road, Suite 500
City: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ. ID NO.: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ19pt-Fls
US-08-232-463-14

```

```

Query Match      2.5%, Score 38.4; DB 1, Length 7216;
Best Local Similarity 1.9%; Pred. No. 0.14;
Matches 6; Conservative 180; Mismatches 126; Indels 0; Gaps 0;

0y 1 aaagtgagttgtgtgtgtgtcgaagtcacaataaagttgtgtgaggtccaactt 60
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

[illegible]

RESULT 4
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9pc-F15
US-08-232-463-14

```

;          TOPOLOGY: linear
;
;          IMMEDIATE SOURCE:
;          CLONE:      PTZqpt-F1s
US-08-232-463-14

```


QY	907	ccacatcaacaattatgatgatgataacccgttgtttgatatttaoctcccaag	966
Db	43310	tactctgatgaaatagaagagatagtataagagattcgggtccctttttaaaaa	43369
QY	967	agtgagacaagagatccctgttgtaggaacacaacgaaagagaattctctggcag	1026
Db	43370	aattcagcttcgatatccccaagaagccaattcattglttaaaaaaattccctgtgct	43429
QY	1027	ggaagctcaagltgtctataaagaaccgggtcccaatgtygatatgtg	1078
Db	43430	tcgaatgacaatgatactttttagatcctcgtgttagcaagtgatattttgtgt	43481

RESULT 8
US-08-972-927-4
; Sequence 4, Application US/08972927
; Patent No. 6166300

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

	Query Match	2.2%	Score 34.8	DB 4	Length 5175
	Best Local Similarity	51.3%	Pred. No. 1.5		
	Matches	81	Conservative	0	Mismatches 77; Indels 0; Gaps 0;
Qy	559	tcagagacaacactttaaacctcaacttgyatattactgctagcaagaatgatcgtgtc	658		
Db	329	TCAGGCTCAAAATGGTTTGATCTATTTTCGGCCCTTTTGGCTCTCTTAATGCTACTGTGGGAC	388		

OY 659 gttgtggaacgttggaacgtagatatttgatattgcgcggagcgagatc 718
 |||||
 Db 389 ctttcttttagatttgctcatagagatctctctttttggatttggatgagacgcggctttcc 448
 |||||
 OY 719 gcatgaagtcctaagtcgtgatttcataccgctctgat 756
 |||||
 Db 449 cctatgagcgcgtttatgttgctccttagagcgttttgct 486
 |||||

RESULT 9
US-08-945-056-4
; Sequence 4, Application US/08945056

APPLICANT: Coupland, George M.
 APPLICANT: Puterill, Joanna J.
 TITLE OF INVENTION: Genetic control of flowering
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon & Vanderhye PC
 STREET: 8th Floor, 1100 No. 6077994th Glebe Road
 CITY: Arlington
 STATE: Virginia
 COUNTRY: United States of America
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/945,056
 FILING DATE: 20-OCT-1997
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/GB95/02561
 FILING DATE: 01-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9422083.7
 FILING DATE: 02-NOV-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Ms Mary J Wilson
 REGISTRATION NUMBER: 32,955
 REFERENCE/DOCKET NUMBER: 620-17
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4201 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Arabidopsis thaliana
 STRAIN: landsberg erecta
 POSITION IN GENOME:
 MAP POSITION: chromosome 5
 US-08-945-056-4

	Query Match	Similarity	2.2%	Score 34.6	DB 3	Length 4201
	Best Local	Similarity	46.5%	Pred. No. 1.6	Mism. 129	Indels 0
	Matches	112	Conservative	0	Matches	129
					Indels	0
					Gaps	0
Qy	138	agactgtgagcttagatagttgtgcgagagcagacagagagagatcgaatcgc	197			
Db	3912	AGAGTAAAGACCCGGCGGTCATGCGCGGTGCCAAGAAGAAATCGAAGCCGAGAGACA	3971			
Qy	198	aatgtcgaagataaacttaaggaaacgagatcgaatagtttggtcttcttgactcactgaataa	257			
Db	3972	AGGGTTCAACACAGATGCTAATGTCACAAACAGAGATATGGGATTTGTCCTTCAATTGATA	4031			
Qy	258	gaactctacaaagtgcacgaataggaattccaggaaggaagaaaccttgcattgcctgtgt	317			

RESULT
 US-08-087-007-4/c
 : Sequence 4, Application US/08087007
 : Patent No. 5705732
 : Patent No. 5705732, 5684223
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Slims, Peter J.
 : APPLICANT: Botwell, Alfred L.M.
 : APPLICANT: Elliott, Eileen A.
 : APPLICANT: Flavell, Richard A.
 : APPLICANT: Madril, Joseph
 : APPLICANT: Rollins, Scott
 : APPLICANT: Bell, Leonard
 : APPLICANT: Squinto, Stephen
 :
 : TITLE OF INVENTION: Universal Donor Cells
 :
 : NUMBER OF SEQUENCES: 4
 :
 : CORRESPONDENCE ADDRES:
 : ADDRESSEE: Kilpatrick & Cody
 : STREET: 1100 Peachtree Street, Suite 28000
 : CITY: Atlanta
 : STATE: Georgia
 : COUNTRY: U.S.
 : ZIP: 30309-4530

```

      FILING DATE: 13-AUG-1996
      CLASSIFICATION: 424
      ATTORNEY/AGENT INFORMATION:
      NAME: Pabst, Patrea L.
      REGISTRATION NUMBER: 31,284
      REFERENCE/DOCKET NUMBER: OMRF 112cipdlv
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (404)-873-8794
      TELEFAX: (404)-873-8795
      INFORMATION FOR SEQ ID NO: 2:
      LENGTH: 315 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      US-08-696-777-2

Query Match          2.1%, Score 33.4; DB 2; Length 315;
Best Local Similarity 51.7%; Pred. No. 0.93;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0.

QY 669 tcttgaacctggagatatttgcgtatatttcgtgaagctgtaggcgcgtatgaagt 728
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 196 TCTTGCACACAGTAGACGTTAGCTCATTTTCCCTCAAGGGGTTGAGCGTGTTGAAMT 137
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 729 tctaagtgtagatttcacccgcgtcgtatattaccgccttgcagttgtgtgatagcaac 788
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 136 TGCATAGCTCAAACTCCCAACACTGTTATATACACTGTGAACCCACTTGTGAATGAGAC 77
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 14
 US-08-483-433-4/c
 ; Sequence 4, Application US/08483433
 ; Patent No. 6100443
 ; GENERAL INFORMATION:
 ; APPLICANT: Sims et al.
 ; TITLE OF INVENTION: Universal Donor Cells
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Patrea L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; STREET: 1201 West Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: U.S.
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/483,433
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/087,007
 ; FILING DATE: July 1, 1993
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/906,394
 ; FILING DATE: June 29, 1992
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pabst, Patrea L.
 ; REGISTRATION NUMBER: 31,284
 ; REFERENCE/DOCKET NUMBER: OMRFL35c4p2 div

2

2

BASE COUNT	167 a	109 c	84 g	122 t
ORIGIN	/clone_11b="Arabidopsis thaliana green siliques Columbia" /issue_type="green siliques" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"			

Query Match	30.1%	Score 470.8;	DB 9;	Length 482;
Best Local Similarity	98.5%	Pred. No. 2.6e-115;		
Matches 475;	Conservative	0;	Mismatches 7;	Indels 0;
				Gaps 0;

Db 482 GGGAGCTTCAGATGTTCTATTAGATACCCGGTCCCATGTGTGATATTTGGTTATCAA 423

Db 422 G T T T T C T T G A C C T C A T T T A A G T T T G T T G C G A T A G G T A T C A G A A G A A G G T T T A 363

Db 362 TCTTTTGGGAAATTTGAAAAGCTTTGCTCTGTTTGTATACAAAGTTATCACAATCATC 303

[illegible]

QY 1266 cgagacgcgactatgcgctggacgtgattaccaagtagcggtctgagctctgag 1325

QY 1326 gaattgatgaattagagtgcgagaagaatgagatatccattctttattgttaattctgat 1389

1386 cagatttactactccctnagacctnagatactctttataacctgttlaacgtccaccctt 1445

D6 122 CATGTTGCTACITCCCIGAGACCIIAGAAATCCICIIATGAGGCCAATGAAAACGCCTCCTTTCCTTG 150

Db 62 GTACCACAGTGATACCCCTTCTGAGATTTCCTTATCTTTAGTTCATACACAAG 3

Db 2 GC 1

RESULT 3
AV519989/C

old Arabidopsis thaliana cDNA clone APD74e07F 3', mRNA sequence

KEYWORDS EST, shale crabs, source

Eukaryota; Viridiplantae; Charophyta; Equisetum; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE A large scale analysis of cDNA in *Arabidopsis thaliana*: General

JOURNAL
DNA Res. 7, 175-180 (2000)

The First Laboratory for Plant Gene Research
varuna DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers

FEATURES

source

1. 486
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="APD74e07F"
/clone.lib="Arabidopsis thaliana aboveground organs two to six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 157 a 101 c 88 g 140 t
ORIGIN

Query Match 28.8%; Score 450; DB 9; Length 486;
Best Local Similarity 96.8%; Pred. No. 9.6e-110;

Matches 459; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 1060 ccaatgctgatatatttggtttatcaagttttctgtgacctcatttaagttctgtcg 1119
DB 486 CCAATGCTGATATTGGTTATCAAGTTTCTGTGACCTCATTTAAGTTCTGTTG 427
OY 1120 atagtaacagaagaaggattatgctgagattgaagaagttgacctctgtttg 1179
DB 426 ATAGTAACAGAGAGAGAAAGTTTATGTTGGATTGAAAAGTTGCTCTGTTTG 367
OY 1180 attacaagaattacacacatcaatcaaatcgtatcaggaagaacagcatgtctgc 1239
DB 366 ATTACAAATGTTATCACATCAATCAATCAAGTCTGTAATCAGGAAACAGCATGCTGTC 307
OY 1240 gatgagaagcagattctgtctgctgagagagagacataatgctgtgagcgtgatt 1299
DB 306 GATGAGAGCAGCATTTCTCTGCTGCGAGAGCGCATATATGCGCTGGAACGTGATT 247
OY 1300 accaagtaagcgtctgagctgtgaagaattgaagaaggagagagagagagagagat 1359
DB 246 ACCAAGTACGCGCTGTAAGCTGTTGAGATTGATGATGAGAGAGAGAGAGAGATGAGAT 187
OY 1360 atccattcttatttgtaattcgtacatgctgtactccctgagagaccttgagatgctc 1419
DB 186 ATCCATTCTTTTATGTATATTCGATCATGTGCTACTCCTCGAGACCTTGATATGCTCTC 127
OY 1420 ttgtgacctgttaacgctccaccttgtaacagtgatatacccttctgagatttgt 1479
DB 126 TTGTAGCCCTGTTAAGCTCACCTGTAACACAGTGTATACCCCTTAAGAAGCCCTGTG 67
OY 1480 ctattctcttgatcaatacagaagcgtgataccgagagcttattgcaagaa 1533
DB 66 CTTATTCTTGTAGTTCAATACAGAAAGTATCTGTGAGCTTTATTGCAAAAA 13

RESULT

4

BM359804 885 bp mRNA linear EST 09-JAN-2002

LOCUS GA_Ea0023003r Gossypium arboreum 7-10 dpa fiber library Gossypium

DEFINITION arDorem cDNA clone GA_Ea0023003r, mRNA sequence.

ACCESSION BM359804

VERSION BM359804.1 GI:18100550

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 885)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)

COMMENT

Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 321
Seq primer: TAATGACTACTACTATAGGG
High quality sequence start: 3
High quality sequence stop: 524.
Location/Qualifiers

FEATURES

source

1. 885
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0023003r"
/clone.lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="fibers isolated from bolls harvested 7-10 dpa"
/note="E. coli"
/note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 258 a 166 c 195 g 266 t
ORIGIN

Query Match 27.5%; Score 430; DB 10; Length 885;
Best Local Similarity 72.3%; Pred. No. 2.7e-104;

Matches 599; Conservative 0; Mismatches 225; Indels 4; Gaps 3;

OY 377 ttactctgtacaatttctctcgagatggtgacctatcagatctgcaatcctatgctgag 436
DB 1 TGACGTTATACCAATGCTCTGAGAGAGGTTTATACCTGTTTGCAGTCTTAATGTTGATC 60
OY 437 aagataagaagagctgctttacacgtaagttggtgctgttaatggaagaccat 496
DB 61 AAGATAGAGATGAGTCTTCTTACACTGTAAGTTGGGCTGCACATTTGATGATACAT 120
OY 497 atgttcgagctgagagagtaaaagataatccagatcattgacgtcaacagtgaagaa 556
DB 121 TTGTGTTGCTGAGGATATATGATGATATCCGTGTAATGATGAGCAATGAGAAAA 180
OY 557 ttataagagctgtgtgtgtatgagatctcattgaagaagaatcagagacaacattta 616
DB 181 TACCAAGAGATTTGTTGGACATGCTGCTTAACAGAAACAGACATCAACCATGA 240
OY 617 aacctcaacttgatcactgctagcaagaagaatcctgtctgtgtgagaa 676
DB 241 AACCATCATTATAGTGTCTGCAAGCAAGATATATCGGTGATGATGATGATGAT 300
OY 677 ctggaatgatatctgatatctgtcgtgagctgagagtcagctatgaagttctaagt 736
DB 301 CTGGAATCTGCAATTCGATTTGTTGCTGAGACCGGGGACCCCAATGAAATTTGAGTC 360
OY 737 tggatttcatcctgctgtatattaccgcttctgctgtgtgtgtatggaacacattat 796
DB 361 TGGATTTCATCTTCAGACATTTATCGAATTCGAATTCGAGTCCGAGTGAACACGTGA 420
OY 797 aatatgctcaatgaagaagcttgtagcgtacgtcagagatcagatcacaatgagcag 856
DB 421 AGATTGGTCAATGAAAGAGTTTGTGACCTTATGTGAGAAATATATTCACATGACAGATC 480
OY 857 atccatcaaatcccccacacaaattgtccaatccctgatttaagcttccatcata 916
DB 481 TTCTTACCAAAATTTCCAAATATATGATGATACAGTACCTGCTGATGATGATGATGAT 540
OY 917 caaatatgtagatgttaacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 976
DB 541 CAATATATGAGCTGATATCTGTGACCTGATTTCAATTTGATCAAAAGAGTGTGAC 600
OY 977 acgagatccgtgtgtggaacacacactgaagaagaattcctgtgagagagctgag 1036

Db 601 ACGAATAGTCTATGGACCAAGATGACGACATCTTCTGAGAACACACACTG 660
 QY 1037 agtctcttaagatacccggttcacatgtgtgatacttgatc-caagttctctgt 1095
 Db 661 ACATCTTTAAAGTACCTCTTCCGGAGTGGACGGTGTTCATTAATAATCTCTTGC 720
 QY 1096 gaccctcatatagctctgtctgataagtaacaggaagaaaggttctgtggat 1155
 Db 721 GAACCTTCATACAAAGCTCTCTGCTATAGS-AAATACAAAGCAAGATTTTGTCTGGAA 779
 QY 1156 ttgaaagttccctcctgttttgatacaaaagtatacaacaatcaa 1203
 Db 780 CTG--AAGGTACCTCTCTGTTATTTGCGACGTTTATGCCAGTCAA 825
 RESULT 5
 BG594585 733 bp mRNA linear EST 12-APR-2001
 LOCUS EST493263 cSTS Solanum tuberosum cDNA clone cSTS7L16 5' sequence,
 DEFINITION mRNA sequence.
 ACCESSION BG594585 GI:13612725
 VERSION BG594585.1
 KEYWORDS EST.
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 733) Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,
 van der Hoeven,R., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
 Generations of ESTs from sprouting potato eyes
 TITLE Unpublished (2000)
 JOURNAL Contact: Cathy Ronning
 COMMENT The Institute for Genomic Research
 For clone info: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cda@resgen.com
 Seq primer: MJ3F-R.
 FEATURES
 source Location/Qualifiers
 1..733
 /organism="Solanum tuberosum"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="cSTS7L16"
 /clone_lib="cSTS"
 /tissue_type="sprouting eyes from tubers"
 /dev_stage="12-14 weeks post harvest"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
 taken from tubers. The tubers were incubated at 26C in the
 dark for 2-3 weeks prior to sprouting. The eyes were
 frozen in liquid nitrogen immediately upon removal from
 tubers."
 BASE COUNT 201 a 127 c 180 g 225 t
 ORIGIN
 Query Match 26.0%; Score 406.4; DB 10; Length 733;
 Best Local Similarity 72.4%; Pred. No. 5.4e-98;
 Matches 527; Conservative 0; Mismatch 201; Indels 0; Gaps 0;

QY 381 tctgtacaattgtctcgagatcgtgccaatcacatgcattgcactcactatgctatgaaga 440
 Db 186 TGTATATGAGTGTCTCGAAGGTGTTTATCTGTGCTGCAAGTCTTATATGATGAAGA 245
 QY 441 taagaagagctgttllacacgtaagtggtcggtgtggttgaatggaaaccataagt 500
 Db 246 TTAAGATGATATCTTTTACAGTGAAGTGGGCTTCGATATGATGAGAGTCAATCTT 305
 QY 501 tgcgctggaaggaataaagataatccagatcatgacgcaacagaggaagatca 560
 Db 306 AGTGCTGTGGAAATGAAGGATTAATGCTGATTAATGCTGGGAAGAGAGATTA 365
 QY 561 taagatcttgggtcagcagatcagtgaaacagacacacacacacacacacacac 620
 Db 366 CAAGACCTTGTGGGACGACGACACTCAATTAATGAATTAAGACTCAACCCCTGAACC 425
 QY 621 tcaactgtgattacgcctagcaagatgaatcgttctgttggatgttgaatgtgaac 680
 Db 426 ATCTCTTGTATATCTGCGACGAAGATGATCTGCTGCTGGAATGCTCATCTG 485
 QY 681 gatagtattgatatctgtcgtgagctgaggtcaatcgtatgaagttcctaagtga 740
 Db 486 AATATGCAATTTGTGATTTGCTGCTGCTGAGGTCAATCGAATGAAGTAACTAGTGGA 545
 QY 741 ttctcactcgtcgtgatatattacgccttgcctgtgtgtgatatgagacacacataa 800
 Db 546 CTTCACATCTGATGATATATATGATGATGCTGCTGCTGATGATGATGATGATGAT 605
 QY 801 atgctcaatgaagagcttltgacgctagcagcagcagcagcagcagcagcagcagc 860
 Db 606 CTGTGCAATGAAGAAATTCGAGCATATGAGAAATCGTTTACTTGGACGATCTTCC 665
 QY 861 atcaaaatcccaaaaatttgcacatccctcgtatttaacagcttccatcaaa 920
 Db 666 TTCAAAGTTTCCACATTAATATGATGATGATGATGATGATGATGATGATGATGAT 725
 QY 921 ttatgtag 928
 Db 726 CTACCTTG 733
 RESULT 6
 BG444266 987 bp mRNA linear EST 15-MAR-2001
 LOCUS GA_Ea0023N03f Gossypium arboreum 7-10 dpa fiber library Gossypium
 DEFINITION arboresum cDNA clone GA_Ea0023N03f, mRNA sequence.
 ACCESSION BG444266 GI:13353918
 VERSION BG444266
 KEYWORDS EST.
 SOURCE Gossypium arboreum.
 ORGANISM Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
 1 (bases 1 to 987)
 REFERENCE 1 (bases 1 to 987)
 AUTHORS Wing,R.A., Prisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
 'D., Wood,T.C., Leslie,A. and Wilkins,T.A.
 TITLE An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 JOURNAL Unpublished (2000)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TTAATGACACTGCTATAGGS
 High quality sequence start: 37
 High quality sequence stop: 727.
 FEATURES
 source Location/Qualifiers
 1..987

```

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/notes="vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      261 a      203 c      263 g      259 t      1 others
ORIGIN

Query Match      23.6%; Score 369.4; DB 10; Length 987;
Best Local Similarity 72.5%; Pred. No. 4.9e-86;
Matches 492; Conservative 0; Mismatches 186; Indels 1; Gaps 1;

QY 400 gatgtgcatatcagcatgcaatcctatctgtaagaagaagaagtcgtttac 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 GAGGTGTCTATAGCTGTTTTCAGCTTATGTGATGAAGATAGATGCTTTCTAC 155

QY 460 acggtgaagtgtggcgctgtgagcgttaatgggaaccatagtgcggctggaggaataaa 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 ACTGTAGTGTGGCCGCAACATGTGATGTACACCATTTGTTGCTGAGGTATTAAT 215

QY 520 ggtataacagatcatgagtcacagtcgaagatcattcaagaagtcgttggtcat 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 GGCATTAATCCGTATATGATGCGACATGAGAAATACACAAGATTTGTTT 275

QY 580 ggaagttcagtgaaacatcagagacacacaccttaaacctcaacttgatctactgct 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 GGTGACTCTATTAAGAAATGAGATCAGGACACGACGAAACCATTAATCCGCTCGA 335

QY 640 agcaagatgaatctgttcgttctgtgaaatcttgaaactggatattgatttgaatt 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 AGCAAGATGATCGGTAGATGTGGAATGTTCACTGAAATCGCATCTGATATTT 395

QY 700 gctggaagtcagtcacatgaattcctaagtcgtgatttcattccgtctgatat 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 GCTGAGCCGGGGACACCGCATAGATTTGATGTGATTTCCATCTTCAACATTT 455

QY 760 taccgcttgcagtcgtgataagacacacacatataatgtaagaagaagttt 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 TATCAATTTGCAAGTTGGGGATGGAACAACGTAAAGATTTGCTCAATGAAGAATTC 515

QY 820 tggagctacgtcagagaatcattcattgactgattgattcaataaattcccaaaaa 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 TGGACTTATGTGAGAAATATTCATTGATGAGCGATCTTACTTCAAAATTTCCACAAAA 575

QY 880 ttgtccaatccctgttatctacagctccattcaataatattgattgattgattgatt 939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 TACGTACAAATTCCTGTCTTCTTACCTTCAGTTCAATTAATGATGACGTATATCGG 635

QY 940 tggatttgatttatctctcaagaagtcgtgcaacagagatcctgttgaggaaaca 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 636 TGGCTTGTGATTTTCTATCTGTCAAGAGATGTGACTACCAAAATCTGCTATGGGAACCA 695

QY 1000 caactgaagaagaattctctcgtgagagagagcttcagatgttcttaataagaccggtt 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 636 AAGATTAAGGAGCA-TCGTCAAGAGAGGAGAACAGCTGACATCTTAAAGAGAGAGCTGCG 754

QY 1060 ccaatgtgtgatttgg 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 755 CCGAATTCGACATTTGTTT 773

RESULT 7
BI424774 585 bp mRNA linear EST 29-NOV-2001
LOCUS BI424774
DEFINITION saH4802.y2 Gm-c1036 glycine max cdna clone GENOME SYSTEMS CLONE
ID: Gm-c1036-4803.5 similar to TR:Q9XF44 Q9XF44
FERTILIZATION-INDEPENDENT ENDOSPERM PROTEIN.; mRNA sequence.

```

```

ACCESSION BI424774
VERSION BI424774.1 GI:15201253
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 585)
REFERENCE Shoemaker,R., Kelm,P., Vodkin,L., Expelling,J., Corryell,V., Khanna
AUTHORS A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Willson,R.
Public Soybean EST Project
Unpublished (1999)
JOURNAL Contact: Shoemaker R/Public Soybean EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccl@resgen.com
High quality sequence stop: 433.
FEATURES
source
1..585
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1036-4803"
/clone_lib="Gm-c1036"
/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/notes="vector: pSPOR1; Site_1: NotI; Site_2: SalI. This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies pSuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dt) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPOR1 vector. The ligated cDNA fragments
were transformed into E.coli Electromax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

BASE COUNT      167 a      97 c      127 g      194 t
ORIGIN

Query Match      23.2%; Score 363.2; DB 10; Length 585;
Best Local Similarity 76.4%; Pred. No. 1.9e-86;
Matches 446; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 643 aaggatgaatcgttcgttcgttggaatgttgaactggagatgtgattgatttgg 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 AAGATGAATCTGTCCGGTATGATGATGTTCACTGCAATTCGATTTTAATTTGGC 60

QY 703 ggaagtcgaagtcacgtcatgaagttcctaagtcgtgatttcacatcgctgataattac 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GGAGCTGGAGGACATCGTAAATGAAGTTTAAGTGTGATTTTCCACCAATATTAATAT 120

QY 763 cgcttgctagtcgtgataagacacacatataaataatgtaagaagaagttgg 822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 CGTATTTGCTAGTTTGGCAGTATGATGATTAATAAATATGCTCATGAAGAGTTCTCG 180

QY 823 acgtacgtcgaagtcattcattcagtcgactgattgattcattcaaatctcccaaat 882

```

Db 181 ACATATGTAGAAAATCATTTACATGACACTGATCTTCAAGTTTCCCAAAATAT 240

Qy 883 gtccatccctcgtatcttaccagctccatccatcaacaattatgtagatgtaaccgttgg 942

Db 241 GTCCAGTTCTCTGTAACAAAGCTTCACTTCAATTAATTAAGTACAGTAATAGGAG 300

Qy 943 ttgtgtatcttctcctcctcaagaagtgtagaacaagagatcctgttggagaccaca 1002

Db 301 CTAGTGTATTTATCTCTCCCAAGAGTGTGACAAATGAATATCTTGGGGAACCTAA 360

Qy 1003 ctgaagaagaattcctcctggcgaagagagcttcaagatgtctcttaagatacccggttcca 1062

Db 361 GTGAAGGAACAACTCCAGGGGAGGCTGTAGTTGACATCTTCAGAAATACCTGTCTT 420

Qy 1063 atgtgtatcttgggttcaagattcttggaccctcaattaaagtcgttgcgata 1122

Db 421 GAGTGTATATCTGTTCTCAAGATTTCTTGTGATTTCCATTTTCACTTACCGCAGTG 480

Qy 1123 ggtaatcaagaagaagttatctgtctggaatttgaagaagtgccctcgttattgatt 1182

Db 481 GGTAAACAGGGAAGGAGATTTTGTGGAATTGACATGACAGTCTCTCTGTACTGTT 540

Qy 1183 acaaaagtatcaacaacatcaatcaaaagtcgttaatcaggccaac 1226

Db 541 GCAAGGTTGTCAATCTCAATCAAAATCTCCATCAGCGAC 584

RESULT 8

BE888279 658 bp mRNA linear EST 30-MAY-2001

LOCUS EST514130 cSTD Solanum tuberosum cDNA clone cSTD9B13 5' sequence,

DEFINITION mRNA sequence.

ACCESSION BE888279

VERSION BE888279.1 GI:14265365

KEYWORDS EST.

SOURCE Solanum tuberosum

ORGANISM Solanum tuberosum

REFERENCE 1 (bases 1 to 658) Bezzerides, J., Ewing, E., Cho, J., Chien, A.,

AUTHORS Bouffard, O., Buehl, C.R., Ronning, C., Tanksley, S. and Baker, B.

TITLE Generations of ESTs from dormant potato tubers

JOURNAL Unpublished (2001)

COMMENT Contact: Cathy Ronning

The Institute for Genomic Research

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com

Seq primer: M13p-R.

FEATURES

source

1. 658

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="cSTD9B13"

/clone_1bp="cSTD"

/tissue_type="dormant tuber"

/dev_stage="one month post-harvest"

/lab_host="SOLR"

Query Match 22.8%; Score 356.8; DB 10; Length 658;

Best Local Similarity 74.7%; Pred. No. 1e-84;

Matches 487; Conservative 0; Mismatches 162; Indels 3; Gaps 3;

Qy 301 cctttagtgcgtgtgttctcaacttcttgatgctgcttcttgatgctgcttcc 360

Db 8 CCATTATATGCGCGTCTTCAACTTCAATGATTTGGCGTACTTAAAGTCTTCTTACT 67

Qy 361 gctgtgtgaatccgagatctctgacaaattgtctcggg-gatgtgcatatcagcatt 419

Db 68 GTGGCGGGAATCGAGTCACTGTATATCATGTCTCGAAGGTTGGTATATGCGTGT 127

Qy 420 gcaatccatgctgtatgaagaagaagaa-gagtcgltttacacggtlaagtgtggcggtg 478

Db 128 GCAGCTTATATGATGATGAAGATTAAGAAATGCAATCCCTTTACACTGTAAAGTTGGCGTCA 187

Qy 479 gcgttaatgggaaccatattgttcgctgagagagtaaaagttataaccgaagtcag 538

Db 188 ATATTGATGGAGTCCATCTTAGTGGCTGTGGAATTAATGAGTATTCGTATTG 247

Qy 539 acgtcaacagtgaaacagatcataagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 598

Db 248 ATGCTGGGAAGAGAGATPACACAGAGCTTGTGTGGGACACGAGACTCAGTAATAAGAA 307

Qy 599 tcaagacacacaccttaaaacctcaactgtgattactgtctagcaagatgaatctgttc 658

Db 308 TTAGGACTCAACCCCTTGAACCAATCTTGTATATCTGACCAAAAGATGATCTGTT 367

Qy 659 gtttgtgaatgttgaactggagatgtatgtattgtattgtctgtgaagtcagtc 718

Db 368 GCTTGTGGAATGTTCAATACGATGATGATGATGATGATGATGATGATGATGATGATG 427

Qy 719 gctatgaagttcctaagtggtgatttcatccgcgtgatattacccgcttgcagtggtg 778

Db 428 GGAATGAAGTACTTGAAGTGGAGCTTCCATCTTCAATATATTCGATTCGATTCGATTCG 487

Qy 779 gtaggaacacactataaataatgtlcaatgaagaagttgtgagctgctgagaagt 838

Db 488 GATGATGATACACTGTTAATGATCTGTCATGAAGAATCTGACATATGATGATGATGAT 547

Qy 839 cattcacatgactgtagatcatcaataatcccccacaattgtccattccctgtat 898

Db 548 CGTTTACTTGGAGGAGTCTTCTTCAAGTTCCTCAATTAATTAATTAATTAATTAATTA 607

Qy 899 ttacagttcatcatcaacaattatgtagatgtgaacggtgtgtgtgtga 950

Db 608 TAATAGCTTCTGTCTCATAC-AACTACGTTGACGTATACATGATGCTGTGTGTA 658

RESULT 9

BE353166 624 bp mRNA linear EST 18-MAY-2001

LOCUS EST340387 tomato mixed elicitor, BRT Lycopersicon esculentum cDNA

DEFINITION clone cLET41E1, mRNA sequence.

ACCESSION BE353166

VERSION BE353166.1 GI:9291142

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 624) Lyman, J., Matern, A.L., Viston, T., Holt, I.E.,

AUTHORS D'Ascenzo, M., He, X., Liang, F., Hansen, T.S., Ronning, C.M., Craven, M.B., Bowman, C.L.,

Nieman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D., Giovannoni, J.J. and Martin, G.B.

TITLE Generation of ESTs from tomato callus (mixed elicitor)

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI

Clemson University Genomics Institute

Clemson University

RESULT 10

910 attcatacaattatgtagattgtaacgcttggtttgtgatttlatcctctcaaaagt 969

Db 374 GTACACTCTACTATCTGACTGACAGATGCGTGTGCTGACTTCATCTATCAAAAGT 433
 QY 970 gtggaacgaagatccgtgtgtggaacacgaactgaagaatctcctgsgaaggga 1029
 Db 434 GTTGACAAATGAATGTGCTGGGAACCAAGCAAAAGAAAGAGAGTCCGGGAGGGA 493
 QY 1030 gcttcagatgtctctaatgaacccggttcacaaatgtgataattgtttatcaagtt 1089
 Db 494 AGCATCATATTTCTTCAAAAGTACCTGCTCCAGAAATGTGACATCTGTTCATCAATTT 553
 QY 1090 tctgttacctccatctaaagttctgttcgaatgaatgaagaagaagttatgtc 1149
 Db 554 TCATGTATTTTCTACTTCATCAATGACTGCTATAGCAATGCTGAAGAAATATACGTG 613
 QY 1150 tggga 1154
 Db 614 TGGGA 618

RESULT 11

BE919873 578 bp mRNA linear EST 02-OCT-2000
 LOCUS BE919873
 DEFINITION EST423642 potato leaves and petioles Solanum tuberosum cDNA clone
 ACCESSION BE919873
 VERSION BE919873
 KEYWORDS EST.
 SOURCE Solanum tuberosum
 ORGANISM potato.
 Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 578)
 van der Hoeven, R.S., Bezzerides, J., Holt, I.E., Liang, F., Cho, J.,
 Utechtack, T., Hansen, C.L., Doan, B., Bougri, O., Buell, C.R., Romning,
 C.M., Fry, W.E., Tanksley, S.D. and Baker, B.
 Generation of ESTs from potato leaves and petioles
 Unpublished (2000)
 Contact: Cathy Romning
 The Institute for Genomic Research
 For clone request: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdna@resgen.com.
 Location/Qualifiers
 1. 578

FEATURES
 source
 1. 578
 /organism="Solanum tuberosum"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="cstb2p18"
 /clone_1db="potato leaves and petioles"
 /tissue-type="leaflets and petioles"
 /dev_stage="8 weeks old plants"
 /lab_host="SODR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; Tissue was supplied by Dr. Fry (Cornell University).
 Leaflets and petioles were isolated from 8 week old
 greenhouse grown plants. The plants were watered and
 fertilized freely. The tissue was immediately frozen in
 liquid nitrogen."
 BASE COUNT 156 a 99 c 149 g 174 t
 ORIGIN

Query Match 20.4%; Score 319; DB 10; Length 578;
 Best Local Similarity 72.2%; Pred. No. 1.3e-74;
 Matches 415; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 201 gtcaagaataccttaggaacgaatgtagtgccttgactcatcgaataaaga 260
 Db 4 GCGGAGATCTACTCTAGATGTGAACCTGTGTGTGATCATTAACGCCCTCAAGAAAA 63
 QY 261 atcgtacaaatgacgaatgtagtcatcaggaaggaagaaccccttgatgtgtgttt 320

Db 64 AGATATAGAGTACGACAGCACTCCAGAAAGCAAGCCCATTTATGCGCTGCTT 123
 QY 321 caacttcctgtagtgcgtcttcttcgactgcttcgtaccgctggtggaatcggatcac 380
 Db 124 CAACTTATTGATTCGGCGTACTTCAATGCTTTGCTACTGTGGGCGGAATGAGTGAC 183
 QY 381 tctgtacaattgtctcggaatggtgcatacagaatcgaatcctatcgtatgaga 440
 Db 184 TGTATACATGTCGTCGACAGGTGCTTATTTGCTGCTCAGCTTATTTGATGAGA 243
 QY 441 taaggaagatcgttttaacacggttaagttggcggtggtggttaagtgaaccatagt 500
 Db 244 TAAAGATGAATCTTTTACCTGATGAGTGGCCCTGCAATATGATGAGGATCCATTCTT 303
 QY 501 tgcggtcggaaggaataaagataacacgaatcgaatcgaatcgaatcgaatcgaatcga 560
 Db 304 AGTGGCTGTGGATTAATATGAGTATTCGTATGATGATGATGATGATGATGATGATGAT 363
 QY 561 taagatcctgtgtggtcagatgagatcagtgagacgaatacgaacacaccccttaaac 620
 Db 364 CAAGAGCTTGTGGGCGCAGGACGACTCAGTAAATGAATTAAGACTACACCTTGAAAC 423
 QY 621 tcaactgtgattactgtcagcaagagatgaatcgttcgttcgttcgttcgttcgttcgt 680
 Db 424 ATCTCTTGTATTAATCTCCAGCAAGATGAATCTTCTGCTGTGGATGTTCTACTGG 483
 QY 681 gatagtatttgaattcgttcgagctggaagcagcagcagcagcagcagcagcagcagcag 740
 Db 484 AATATGCAATTTTGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
 QY 741 ttctacccgctcgtatcttaacgcttgcagtt 775
 Db 544 CTTCATCTCTACTGATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 578

RESULT 12

BE256342 773 bp mRNA linear EST 22-OCT-2001
 LOCUS BE256342
 DEFINITION HVSMEF0009E24f Hordeum vulgare seedling root EST library HVCNDA0007
 (Etisolated and unstressed) Hordeum vulgare cDNA clone
 HVSMEF0009E24f, mRNA sequence.
 BE256342
 VERSION BE256342
 KEYWORDS EST.
 SOURCE Hordeum vulgare
 ORGANISM barley.
 Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 1 (bases 1 to 773)
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
 Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton
 R.D., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex unstressed seedling root cDNA library
 Unpublished (2001)
 On Nov 16, 2000 this sequence version replaced g1:1185371.
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total bp bases = 545
 Seq primer: AATTACCTTCACTAAGG
 High quality sequence stop: 695.
 Location/Qualifiers
 1. 773

FEATURES
 source
 1. 773
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMEF0009E24f"

Qy	911	ttctacaataatgtagatgtagaacctggtgtttgtgtatcttcatcctccaaagggtg	970
Db	181	ttccattacacactgactgactgttaacagatggccttggtgattttatcttgcacaaagctg	240
Qy	971	tgagcaacgagatccctgtgtgtgggaaccacaacactgaaagagatctcctctgcgagggag	1030
Db	241	ttgcacatggaattctattatgggagaccacaagatgaaagaacattcctctgcgagggcca	300
Qy	1031	ctcagaatgctctatlaagatacccggttccaatgtgtgatactttgatttcaagttc	1090
Db	301	ccagagagctctctcccaaaagtatctctgtgcccggagtgatgattgtgtttttaaagcttt	360
Qy	1091	cttgtgacctcatttaagtctgtgttgatagataacacgaagaaaggtttatgctt	1150
Db	361	catgttgattttacactacacaaaacagacgttatatgggaattatgcaaaagcgaagctctttgtgctt	420
Qy	1151	ggagattgaaaaagtgccctcctctgttttggattacaagaattcacacacatcaatcaagaat	1210
Db	421	gggaagtgtaaaacaaagaccgccgaactttgattgcaaaagctgtctatgttcaacttaaac	480
Qy	1211	ctgtaatcaaggcaaacagacatgctctgcgatgtagaagcagatctctgtctgcggagg	1270
Db	481	aaccattatgacacagactcccatgctcttttgatggaagcaccatactgttgcgttgtaag	540
Qy	1271	acggagcatatagcgctgcgtggaacgtgattaca	1303
Db	541	atgggactatataaggcgtgggaatggtaacca	573

RESULT	14
AV914549	
LOCUS	
DEFINITION	AV914549 608 bp mRNA linear EST 18-JAN-2002
ACCESSION	AV914549 K. Sato unpublished cDNA library, cv. Haruna Nijo
VERSION	gemination shoots Hordeum vulgare subsp. vulgare cDNA clone
SOURCE	bag608 5', mRNA sequence.
KEYWORDS	AV914549 GI:18210326
ORGANISM	EST.
	Hordeum vulgare subsp. vulgare
	Hordeum vulgare subsp. vulgare

REFERENCE 1 (bases 1 to 608)
AUTHORS Sato, R., Saitoh, D. and Takeda, K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL unpublished (2002)
COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel.: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES	location/Qualifiers
source	1. .608

BASE COUNT	ORIGIN
153 a	135 c 142 g 177 t 1 others

Query Match	19.8%;	Score 309.4;	DB 9;	Length 608;
Best Local Similarity	69.2%;	Pred. No. 5,1e-72;		
Matches 421;	Conservative	0;	Mismatches 187;	Indels 0;
				Gaps 0;

QY 358 accgcctgctgggaatcgtgattactctgtcaaatctctcgtgagatgctgcataacgca 417

Db	1	ACCGTCGGCGGCATTCCTGTGACGACCTACCGTCCCTCCCGCGGCTAACTTGGCGTTT	60
OY	418	ttggaatccctatgctgtgttgaagaagaagtcgcgtttcaacggyaaattgagcgtt	477
Db	61	CTTAAAGCATTCATTGTAGTGGGACCAATGCTCAGTCACTTTCACCCCTGACCTGGCGTTGT	120
OY	478	ggcgctaaatggyaaaccatattgttcggtctlgagagtaaaaglataatccgaagtcatt	537
Db	121	GACCTTGACGGCACACCACTGCTGTGTGGACAGCAAGCAATCGCGGTCAATTCGGGCTATC	180
OY	538	gaagtcacacgttgaacagatccatcaagaagtcctgttggtgtaatgagattcaagtaacgaa	597
Db	181	AACGTGTGCCACCGAAGTGTGTTTAAGACTTTCTTGCCATGCGTATTCATATTAATGAG	240
OY	598	atcagagacacacactttaaaccctcaacttltgattcaactcgttagcaagatgatcgtt	657
Db	241	ATAGAATCTCAATCAACATTGAACCTTCATCTTCACTTCGCAMCAAGAGAGCACTCTGTT	300
OY	658	cgtttgtggaatgttgaacacgtggagatatgtatttggtaatttggtagagctgagagtcatt	717
Db	301	AGGTTATGTGAATGTCACATACAGGGATCTGCATTTTGATTTTGGCTGGAGGAGAGGTCAC	360
OY	718	cgctatgaagttcctaagtgltgatttccatccgcctcgtatattcaacgccttgcagattgt	777
Db	361	CGTATATGAATGATTTAGTGTGACTGCCACCTCTGTGATTCFACCGAATGCGCAGTGTG	420
OY	778	ggtatgysaacocactataaataatgltcaatgaaagagtttggagtgtaagtcgagaag	837
Db	421	GGCATGAGTATTACTGTTAAATCTGGTCAATGAAAGAAATTTTGGCCATACGTGAGAA	480
OY	838	tcaattcacatgactgatatccatcaaaattccccaacaaaatttgcacattccctcgtla	897
Db	481	TGCTTTAATGACGTGACCTTCATCAAAATTTCCAAAATAATGTCGAATTTGCCACTT	540
OY	898	tttaacgcttccatccatcaacaatbatgtagatttgaacggttggttgggtgatttcatc	957
Db	541	ATGACTTCGCGGTGCATTTCATATGTTAGCTGTACTAGGTGCTTGGTACTTTCATC	600
OY	958	cctctcaaa	965
Db	601	CTGTGCGAA	608

RESULT	15			
LOCUS	BG132016			
DEFINITION	BG132016	566 bp	mrna	linear
ACCSSION	EST464908	tomato crown gall	Ycopersicon	esculentum
VERSION	CTO6E61.1	5' sequence,	mrna	sequence.
KEYWORDS	BG132016			
	EST.	GI:12632204		
				EST 31-JAN-2001
				CDNA clone

ORGANISM *Lycopersicon esculentum*
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;

Eubaijota; Vitaceae; Convolvulaceae; Solanaceae; Scrophulariaceae; Asteraceae; Magnoliophyta; eudicotyledons; core eudicots; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae I: Solanales; Solanales; Solanaceae; Solanum; Asteridae I: Solanales; Solanales; Solanaceae; Solanum.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 566)	van der Hoeven, R., Sun, H., Cho, J., Uterback, T., Hansen, C., Romling, C. and Tanksley, S.	Generation of ESTs from tomato crown gall tissue	Unpublished (2001)	Contact: CUGI

Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>

```

FEATURES
  source      location/Qualifiers
1..566
/organism="Lycomorpha esculentum"
/cultivar="T7496"
/db_xref="taxon:4081"

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 01:13:56 ; Search time 88.61 Seconds

(without alignments)
720.405 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973
Sequence: 1 MSKITIGNESIVGSLRPSNR.....STIACCEDEGTWRMDVITK 369

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp rodent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.rv1rus:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1973	100.0	369	10	Q9XF44
2	1565	79.3	294	10	Q9L747
3	727.5	36.9	441	4	O00149
4	727	36.8	427	4	O9UNY7
5	727	36.8	441	11	O921E6
6	727	36.8	517	4	O75530
7	727	36.8	535	11	P97462
8	718	36.4	462	5	O9VKD5
9	717	36.3	426	13	O90YL5
10	715	36.2	425	5	O24338
11	708	35.9	425	5	O26458
12	703	35.6	428	5	O16023
13	690	35.0	412	5	O16021
14	689.5	34.9	437	5	O16022
15	370	18.8	190	5	O95SM6
16	300	15.2	459	5	O16187

17	300	15.2	536	5	O9GXS1	O9GXS1 caenorhabd
18	211.5	10.7	665	5	O9GNM6	O9GNM6 caenorhabd
19	200	10.1	430	5	O9B54	O9B54 heterodera
20	197.5	10.0	913	10	O9CAAO	O9CAAO arabidopsis
21	197.5	10.0	920	10	O9SSD5	O9SSD5 arabidopsis
22	194.5	9.9	911	10	O9LW87	O9LW87 arabidopsis
23	194.5	9.9	926	10	O9C827	O9C827 arabidopsis
24	193.5	9.8	252	11	O922C7	O922C7 mus musculu
25	188	9.5	860	5	O9VJ20	O9VJ20 drosophila
26	187.5	9.5	421	4	O9UBH9	O9UBH9 homo sapien
27	187.5	9.5	421	4	O9Y617	O9Y617 homo sapien
28	186	9.4	421	11	O54927	O54927 mus musculu
29	185.5	9.4	569	11	O92159	O92159 mus musculu
30	183.5	9.3	421	4	O9UNM6	O9UNM6 homo sapien
31	183.5	9.3	1326	5	O9VZF4	O9VZF4 drosophila
32	182	9.2	377	10	O9SW94	O9SW94 pisum sativ
33	181.5	9.2	510	5	O44382	O44382 drosophila
34	181.5	9.2	510	5	O9VDE3	O9VDE3 drosophila
35	180.5	9.1	553	4	O9NMX6	O9NMX6 homo sapien
36	180.5	9.1	561	4	O96R12	O96R12 homo sapien
37	180.5	9.1	589	4	O96LE0	O96LE0 homo sapien
38	180.5	9.1	627	4	O96A16	O96A16 homo sapien
39	180.5	9.1	707	4	O969H0	O969H0 homo sapien
40	178.5	9.0	569	11	O9QU15	O9QU15 mus musculu
41	177	9.0	655	4	O9BVA0	O9BVA0 homo sapien
42	177	9.0	657	11	O9CWM2	O9CWM2 mus musculu
43	176.5	8.9	335	5	O9BNG5	O9BNG5 aedes aegypt
44	176.5	8.9	335	5	O9BMG4	O9BMG4 aedes aegypt
45	176.5	8.9	563	11	O923H0	O923H0 mus musculu

ALIGNMENTS

RESULT 1

O9XF44 PRELIMINARY: PRT: 369 AA.

AC O9XF44; 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE FERTILIZATION-INDEPENDENT ENDOSPERM PROTEIN.

GN FIE

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. LANDSBERG ERECTA;

RX MEDLINE=99172055; PubMed=10072400;

RA Ohad N., Yadegari R., Margossian L., Hannon M., Michaeli D.,

RA Harada J.J., Goldberg R.B., Fischer R.L.;

RT "Mutations in FIE, a WD polycarb group gene, allow endosperm development without fertilization."

RL Plant Cell 11:407-416(1999).

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC EMBL: AF129516; AAD23584.1; .

DR InterPro: IPR001680; WD40.

DR Pfam: PR00400; WD40; 7.

DR PRINTS: PR00320; GPROTEINRPT.

DR SMART: SM00320; WD40; 3.

DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.

DR PROSITE: PS00682; WD_REPEATS_2; 2.

DR PROSITE: PS50294; WD_REPEATS_REGION; 1.

KM Repeat: WD repeat.

SO SEQUENCE 369 AA; 41257 MW; EDF0989C02B4FDDA CRC64;

Query Match 100.0%; Score 1973; DB 10; Length 369;
Best local Similarity 100.0%; Pred. No. 1.7e-176;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	MSKITLGNESITVSGLTLPENKKSYKVTNRIOGKRPLYAVVFNFLDAREPDEVVTAAGNRI	60
Dd	1	MSKITLGNESITVSGLTLPENKKSYKVTNRIOGKRPLYAVVFNFLDAREPDEVVTAAGNRI	60
OY	61	TLVNLGCGAISALOSVADEDEKEESFYTVSMACGVNGNPVYAAGVKIIIRYIDVNSETI	120
Dd	61	TLVNLGCGAISALOSVADEDEKEESFYTVSMACGVNGNPVYAAGVKIIIRYIDVNSETI	120
OY	121	HKSLVGHSDSVNEIRTOPLKPOLVTTASKDESRYLMNVTGCIILIFAGAGHREVLISV	180
Dd	121	HKSLVGHSDSVNEIRTOPLKPOLVTTASKDESRYLMNVTGCIILIFAGAGHREVLISV	180
OY	181	DEHPEDIRFASCGMDTTIKITMSMKEFVTYEEKSFTTWDDPSKFPETKVOEPVFTASIIHT	240
Dd	181	DEHPEDIRFASCGMDTTIKITMSMKEFVTYEEKSFTTWDDPSKFPETKVOEPVFTASIIHT	240
OY	241	NYVDGNRMFGDFILISKSVDNELLMEPOLKENSPEGASDVLLARYVPVMDIWFITKFSCD	300
Dd	241	NYVDGNRMFGDFILISKSVDNELLMEPOLKENSPEGASDVLLARYVPVMDIWFITKFSCD	300
OY	301	LHLSSVAIGNOEGKRYVNDLKSCEPVLLTTLKLSHNSKSVIRQTASVSGSTLLACCDEGT	360
Dd	301	LHLSSVAIGNOEGKRYVNDLKSCEPVLLTTLKLSHNSKSVIRQTASVSGSTLLACCDEGT	360
OY	361	IWRMDVITK 369	
Dd	361	IWRMDVITK 369	
RESULT	2		
O9LT47			
ID	O9LT47	PRELIMINARY;	PRT; 294 AA.
AC	O9LT47		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	FERTILIZATION-INDEPENDENT ENDOSPERM PROTEIN-LIKE.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
RA	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;		
RL	Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
RX	MEDLINE=20277480; PubMed=10819329;		
RA	Nakamura Y.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence		
RT	features of the regions of 4,504,864 bp covered by sixty P1 and TAC		
RT	clones."		
RC	DNA Res. 7:131-135(2000).		
CC	-I- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).		
DR	EMBL; AB025629; BAB02481.1; -		
DR	InterPro; IPR001580; WD40.		
DR	Pfam; PF00400; WD40; 6.		
DR	PRINTS; PR00320; GPROTEINBRPT.		
DR	SMART; SM00320; WD40; 3.		
DR	PROSITE; PS00678; WD_REPEATS.1; UNKNOWN_1.		
DR	PROSITE; PS00678; WD_REPEATS.2; 2.		
DR	PROSITE; PS0294; WD_REPEATS_REGION; 1.		
DR	Repeat; WD repeat.		
KW	SEQUENCE 294 AA; 33024 MW; A9A2A1DC83A8B4CD CRC64;		
Query Match	79.3%;	Score 1565;	DB 10; Length 294;
Best Local Similarity	99.7%;	Pred. No.2.2e-138;	
Matches 289; Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;

Qy	80	EDKESSEFTVSMAGVGNPNPVAAGVYKGIIRIVDVSSETHKSLVGHGVSVEIRPPL	139
Db	5	EOKESSEFTVSMAGVGNPNPVAAGVYKGIIRIVDVSSETHKSLVGHGVSVEIRPPL	64
Qy	140	KPOLYITASKDESVRLMNVETGICILIFAGAGGHRVEYLSVDHPSDIYRPAASGMPTI	199
Db	65	KPOLYITASKDESVRLMNVETGICILIFAGAGGHRVEYLSVDHPSDIYRPAASGMPTI	124
Qy	200	KIMSKKEFWTVVEKSEFTWTDPSKEPTKEVQFPVETASIHINYVDCKRMKGDFILSKSYD	259
Db	125	KIMSKKEFWTVVEKSEFTWTDPSKEPTKEVQFPVETASIHINYVDCKRMKGDFILSKSYD	184
Qy	260	NEILLIMEPOLKENSPGEGASVLLRTPVPMCDIMFIFECSDLHLSVAIGNOBSKYVMD	319
Db	185	NEILLIMEPOLKENSPGEGASVLLRTPVPMCDIMFIFECSDLHLSVAIGNOBSKYVMD	244
Qy	320	LKSCPPVILITKLSHNOSKSVIROTAMASVDSSTILLACEDGTIRMDVIYTK	369
Db	245	LKSCPPVILITKLSHNOSKSVIROTAMASVDSSTILLACEDGTIRMDVIYTK	294

ID	000149		PRELIMINARY;	PRT:	441 AA.
AC	000149,				
DT	01-JUL-1997	(TREMBLrel_04_Created)			
DT	01-NOV-1998	(TREMBLrel_08_Last sequence update)			
DT	01-DEC-2001	(TREMBLrel_19_Last annotation update)			
DE	EMBRYONIC ECTODERM DEVELOPMENT PROTEIN HOMOLOG.				
CN	EED.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OX	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Peyravi R., Hong S.S., Gay B., Dupuy D'Angeac A., Selig L.,				
RA	Benichou S., Benarous R., Boulanger P.;				
RT	"Heed, the product of the human homolog of the murine eed gene, binds				
RL	to the matrix protein of HIV-1.";				
J.	J. Biol. Chem. 0:0-0(1998).				
RN	[2]				
RP	SEQUENCE OF 1-400 FROM N.A.				
RC	TISSUE=SPLEEN;				
RX	MEDLINE=99098909; PubMed=9880543;				
RA	Peyravi R., Hong S.S., Gay B., d'Angeac A.D., Selig L., Benichou S.,				
RA	Benarous R., Boulanger P.;				
KHPD,	the product of the human homolog of the murine eed gene, binds				
to the,	matrix protein of HIV-1,";				
J.	J. Biol. Chem. 274:1635-1645(1999).				
CC	-E- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).				
DR	EMBL; U09651; AAD08714..1; "- "				
DR	EMBL; AF099032; AAD08815..1; "- "				
DR	InterPro; IPR004048; RCCL.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF00400; WD40; 5.				
DR	PRINTS; PR00320; GPROTEINRPT.				
DR	SMART; SM00320; WD40; 3.				
DR	PROSITE; PS00626; RCCL_2; UNKNOWN_1.				
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.				
DR	PROSITE; PSS0082; WD_REPEATS_2; 2.				
DR	PROSITE; PSS00294; WD_REPEATS_REGION; 1.				
KW	Repeat; WD repeat.				
SO	SEQUENCE 441 AA; 50155 MW; 2EA5A5BEFEA56B0 CRC64;				

Query Match 36.9%; Score 727.5; DB 4; Length 441;
 Best Local Similarity 41.2%; Pred. No. 1.3e-59;
 Matches 153; Conservative 74; Mismatches 115; Indels 29; Gaps 14;

GY 18 SNKK---SYKVNRRIOGS-KRKPLAVVFNEFLDAFPFD--VPFTAGGNRTLYNCNGDGAI 71
 1-11 1-11 1-11 :111: | 11: :: | 11 | 11|1111 | 1 |

```

Db 74 SSKKCYSEFKVNSLKEDHNOPLFGVQFMW-HSKEGDPLVFATVGSNRYLYECHSGDEL 132
Qy 72 SALQSTADDEKESFTYVSMACGVN-GNPYVAAGVKGIIRIVDENSETHKSLVGHGDS 130
Db 133 RLQSQVADADENEFYTCAMTYDSNTSHPLAVAGSGRIIRINIPITMOCIKHYVGHNA 192
Qy 131 VNEIRTOPLPOLVITASDESIVRLNVEGICILIFAGAGHREVLVSDFHPSDIY-- 188
Db 193 INELKFNPDPLILSVSDHAIKRLMINIOTDLVAIFGVEGHRDVLADY---DLGE 249
Qy 189 RPAACGMDTTIKIW--SMKEFTYVEKSFPTWDDPSK---FPTKVOFPVF-TASIHYN 241
Db 250 KIMSCGMDHSLKLMIRNSKRMMAIKESYD--NPNKTRNPFISQKIHPPDFSTRDIHRN 307
Qy 242 YVDCNRMEDFLLSKSVNDEILMEP-QLKEN-----SPEGASDYLRIYPKMDIWFIK 296
Db 308 YVDCVFMIDLLSKSCENAIYCMKPGKMKDDIKIPSESNVTILGRDYSCDDIWMYR 367
Qy 297 FSCDHLSSVAIGNOGKYYVWDLKSPP--VLITKLSHNSKSVYROTAMSVDSSTIIA 354
Db 368 FEMDFQKMLALGNQYKLYVMDLEVEDPHKAKCTTLTHHKCGAIIROTSPFSKDSIIIA 427
Qy 355 CCEGDTIWRMD 365
Db 428 VCDASITWRMD 438

```

```

RESULT 4
QUNY7 PRELIMINARY; PRT; 427 AA.
AC 09UNY7:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE WAIT-1.
GN WAIT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=98438520; PubMed=9765275;
RA Riezler M., Bittner M., Kolanus W., Schuster A., Holzman B.;
RT "The human WD repeat protein WAIT-1 specifically interacts with the
RT cytoplasmic tails of beta7-integrins."
RL J. Biol. Chem. 273:27459-27466(1998).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF078933; AAC68675.1;
DR InterPro: IPR004048; RCCL.
DR InterPro: IPR001680; WD40.
DR PIRam: PF00400; WD40; 5.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS00682; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 427 AA; 48779 MW; F2A4FE5A09CF9ACD CRC64;

```

Query Match 36.8%; Score 727; DB 4; Length 427;
 Best Local Similarity 41.3%; Pred. No. 1.3e-59;
 Matches 151; Conservative 70; Mismatches 119; Indels 26; Gaps 12;

```

Qy 20 KSKYKYNRIQEG-KKPLAAYVNFIDARFED--VFVTAGNRIITLYNCLGDAISALOS 76
Db 65 KSKFKVNSLKEDHNOPLFGVQFMW-HSKEGDPLVFATVGSNRYLYECHSGGEIRLLOS 123
Qy 77 YADEKDESEFTYVSMACGVN-GNPYVAAGVKGIIRIVDENSETHKSLVGHGDSYNEIR 135
Db 124 YVDADADENEFYTCAMTYDSNTSHPLAVAGSGRIIRINIPITMOCIKHYVGHNAINELK 183

```

```

Qy 136 TOPLPOLVITASKDESIVRLNVEGICILIFAGAGHREVLVSDFHPSDIY--RFASC 193
Db 184 FHRDPNLLSLVSKDHAIKRLMINIOTDLVAIFGVEGHRDVLADY---DLGEIMSC 240
Qy 194 GMDTTIKIW--SMKEFTYVEKSFPTWDDPSK---FPTKVOFPVF-TASIHYNVDCN 246
Db 241 GMDHSLKLMIRNSKRMMAIKESYD--NPNKTRNPFISQKIHPPDFSTRDIHRNVDCV 298
Qy 247 RWFGDPLILSKSVNDEILMEPOLKEN-----SPEGASDYLRIYPKMDIWFIKSCL 301
Db 299 RWFGLDILSKSCENAIYCMKPGKMKDDIKIPSESNVTILGRDYSCDDIWMYRSMDF 358
Qy 302 HLSVAIGNOGKYYVWDLKSPP--VLITKLSHNSKSVYROTAMSVDSSTIIACCEDG 359
Db 359 WQKMLALGNQYKLYVMDLEVEDPHKAKCTTLTHHKCGAIIROTSPFSKDSIIIAVCDDA 418
Qy 360 TIWRMD 365
Db 419 SIWRMD 424

```

```

RESULT 5
QUNY7 PRELIMINARY; PRT; 441 AA.
AC 0921E6:
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE SIMILAR TO EMBRYONIC ECTODERM DEVELOPMENT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 11
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012966; AAH12966.1;
SQ SEQUENCE 441 AA; 50197 MW; D2E0A5BA27C0499A CRC64;

```

Query Match 36.8%; Score 727; DB 11; Length 441;
 Best Local Similarity 41.3%; Pred. No. 1.4e-59;
 Matches 151; Conservative 70; Mismatches 119; Indels 26; Gaps 12;

```

Qy 20 KSKYKYNRIQEG-KKPLAAYVNFIDARFED--VFVTAGNRIITLYNCLGDAISALOS 76
Db 79 KSKFKVNSLKEDHNOPLFGVQFMW-HSKEGDPLVFATVGSNRYLYECHSGGEIRLLOS 137
Qy 77 YADEKDESEFTYVSMACGVN-GNPYVAAGVKGIIRIVDENSETHKSLVGHGDSYNEIR 135
Db 138 YVDADADENEFYTCAMTYDSNTSHPLAVAGSGRIIRINIPITMOCIKHYVGHNAINELK 197
Qy 136 TOPLPOLVITASKDESIVRLNVEGICILIFAGAGHREVLVSDFHPSDIY--RFASC 193
Db 198 FHRDPNLLSLVSKDHAIKRLMINIOTDLVAIFGVEGHRDVLADY---DLGEIMSC 254
Qy 194 GMDTTIKIW--SMKEFTYVEKSFPTWDDPSK---FPTKVOFPVF-TASIHYNVDCN 246
Db 255 GMDHSLKLMIRNSKRMMAIKESYD--NPNKTRNPFISQKIHPPDFSTRDIHRNVDCV 312
Qy 247 RWFGDPLILSKSVNDEILMEPOLKEN-----SPEGASDYLRIYPKMDIWFIKSCL 301
Db 313 RWFGLDILSKSCENAIYCMKPGKMKDDIKIPSESNVTILGRDYSCDDIWMYRSMDF 372
Qy 302 HLSVAIGNOGKYYVWDLKSPP--VLITKLSHNSKSVYROTAMSVDSSTIIACCEDG 359
Db 373 WQKMLALGNQYKLYVMDLEVEDPHKAKCTTLTHHKCGAIIROTSPFSKDSIIIAVCDDA 432
Qy 360 TIWRMD 365
Db 433 SIWRMD 438

```

```

RESULT 6
ID 075530 PRELIMINARY: PRT: 517 AA.
AC 075530;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE EMBRYONIC ECTODERM DEVELOPMENT PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98252946; PubMed-9584199;
RA Sewalt R.G.A.B., van der Vleg J., Hendrix T., van Driel R., Otte A.P.;
RA den Blaauwen J.L., Salijn D.P.E.,
RT "Characterization of interactions between the mammalian polycomb-group
RT proteins Eml1/Ezh2 and EED suggests the existence of different
RT mammalian polycomb-group protein complexes."
RL Mol. Cell. Biol. 18:3586-3595(1998).
RN [2]
RP SEQUENCE OF 77-517 FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-99026128; PubMed-9806832;
RA Schumacher A., Lichtarge O., Schwartz S., Magnuson T.;
RT "The murine Polycomb-group gene eed and its human orthologue:
RT functional implications of evolutionary conservation."
RL Genomics 54:79-88(1998).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF070418; AAC23685.1; -;
DR EMBL: AF080227; AAC95144.1; -;
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT NON_TER
SQ SEQUENCE 517 AA; 57907 MW; 5711CAC7630CCB7A CRC64;

```

```

Query Match 36.88; Score 727; DB 4; Length 517;
Best Local Similarity 41.38; Pred. No. 1.8e-59;
Matches 151; Conservative 70; Mismatches 119; Indels 26; Gaps 12;

OY 20 KSKYKVTNRIOEG-KKPLAVAVFNLDAFPD--VFVTAGGNRIITLYNCLDGAISALQS 76
DB 155 KYSEFCVNSLKDHNQPLFVGQFNW-HSKEDGDLVFAVGSNRYVLYECHSGOGEIRLLQS 213
OY 77 YADEDKESFPTVSNACGVN-GNPVVAAGVKGITRIVDVNSETIHKSLVGHGDSVNEIR 135
DB 214 YVDADADENFTCAVTSNTSHPLAAGSRGIIIRIINPTMOCIKHYVGHGAINELK 273
OY 136 TOPLPOLVITASKDESRLNWNVEIGICILIFAGAGHRYEVLSDVFPSPDIY--RFASC 193
DB 274 FHRPDPNLLISVSKDHALRINVIOTDLVALIFGVGHRDEVLSDY--DLGKIMSC 330
OY 194 GMDTITIKT--SMKEFWTVKESFTWDDPSK---PFTKVOFPVF--TASIHRTVYDCN 246
DB 331 GMDHSLKLMRINSKRMMAIKESYDY--NPKNTRNPFISQKHFPDEFSTROIHRHYVDCV 388
OY 247 RWFGEFLLSKSVNDLILMEPOLKEN-----SPGEGASDVLLRYVPVPCDITWTFISCDL 301
DB 389 RNLGDLILSKSCENAIYVCMKPGKMEDIDIKIKPSESNTIIGREFYSOCDIWMFRFSMDF 448
OY 302 HLLSVAIGNOEKRYVWDLKSCP--VLITKLSHQSKSVIRQTAMSYDGSITILACEDG 359
DB 449 WOKMLATGNQGVKLYVMDLEVEDPHKACTTLTHHKGCAIROTSPSRDSTILAVCDA 508

```

```

OY 360 TIRWRD 365
DB 509 SIWRWD 514

RESULT 7
ID P97462 PRELIMINARY: PRT: 535 AA.
AC P97462;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-JAN-1998 (TREMBlrel. 05, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE EMBRYONIC ECTODERM DEVELOPMENT (EMBRYONIC ECTODERM DEVELOPMENT
DE PROTEIN).
GN EED OR EED.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Denisenko O., Bomsztyk K.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 95-535 FROM N.A.
RC STRAIN=SWISS WEBSTER;
RX MEDLINE-96353973; PubMed-8717038;
RA Schumacher A., Faust C., Magnuson T.;
RT "Role of leptin in the neuroendocrine response to fasting."
RL Nature 382:250-253(1996).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: U97675; AAC53302.1; -;
DR EMBL: U78103; AAB38319.1; -;
DR MGD: MGI:95286; Eed.
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR PRINTS: PR00320; GPROTEINRPT.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 535 AA; 59997 MW; 7FF1216AE6EE497A CRC64;

```

```

Query Match 36.88; Score 727; DB 11; Length 535;
Best Local Similarity 41.38; Pred. No. 1.9e-59;
Matches 151; Conservative 70; Mismatches 119; Indels 26; Gaps 12;

OY 20 KSKYKVTNRIOEG-KKPLAVAVFNLDAFPD--VFVTAGGNRIITLYNCLDGAISALQS 76
DB 173 KYSEFCVNSLKDHNQPLFVGQFNW-HSKEDGDLVFAVGSNRYVLYECHSGOGEIRLLQS 231
OY 77 YADEDKESFPTVSNACGVN-GNPVVAAGVKGITRIVDVNSETIHKSLVGHGDSVNEIR 135
DB 232 YVDADADENFTCAVTSNTSHPLAAGSRGIIIRIINPTMOCIKHYVGHGAINELK 291
OY 136 TOPLPOLVITASKDESRLNWNVEIGICILIFAGAGHRYEVLSDVFPSPDIY--RFASC 193
DB 292 FHRPDPNLLISVSKDHALRINVIOTDLVALIFGVGHRDEVLSDY--DLGKIMSC 348
OY 194 GMDTITIKT--SMKEFWTVKESFTWDDPSK---PFTKVOFPVF--TASIHRTVYDCN 246
DB 349 GMDHSLKLMRINSKRMMAIKESYDY--NPKNTRNPFISQKHFPDEFSTROIHRHYVDCV 406
OY 247 RWFGEFLLSKSVNDLILMEPOLKEN-----SPGEGASDVLLRYVPVPCDITWTFISCDL 301
DB 407 RNLGDLILSKSCENAIYVCMKPGKMEDIDIKIKPSESNTIIGREFYSOCDIWMFRFSMDF 466
OY 302 HLLSVAIGNOEKRYVWDLKSCP--VLITKLSHQSKSVIRQTAMSYDGSITILACEDG 359

```



```

DB 467 WCKMLALGNQVCKLYWDELEVEDPHKACTTTLTHKCGAIRTSPRSSSLIAVCCDA 526
QY 360 TWRWD 365
DB 527 TWRWD 532

RESULT 8
Q9VKD5 PRELIMINARY; PRT; 462 AA.
AC Q9VKD5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE CG5202 PROTEIN.
GN CG5202.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxId=7227;
RX MEDLINE=20196006; PubMed=10731132;
RC STRAIN-BERKELEY;
RA Adams M.D., Cealiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amandlis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Gurgeon R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abdyan A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Boriva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Mostreli A.,
RA Mount S.M., Moy W., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkiamos I., Simpson M.C., Skupski M.P., Smith T.,
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter K., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC EMBL: AB003634; AAF53141.1; -
DR EMBL: AB003634; AAF53141.1; -
DR FlyBase: FBgn0032391; CG5202.
DR InterPro: IPR000408; KCC1.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 6.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RCL1_2; UNKNOWN_1.

```

```

DR PROSITE: PS00678; WD_REPEATS_1; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 462 AA; 52435 MW; A3778A18186158DC CRC64;

```

```

Query Match 36.4%; Score 718; DB 5; Length 462;
Best Local Similarity 39.5%; Pred. No. 1e-58;
Matches 155; Conservative 72; Mismatches 121; Indels 44; Gaps 14;

```

```

QY 14 SLTPSNKSKYKTYNRIOEGK-KPLY-----AVFN-FLDARFDFVFA 55
DB 72 SKSPNTRKLNRLCRKAPRMVPLTKYSSHVEDNHQIFGQFNPFLDRGPQVFATY 131
QY 56 GGNRTFLYNC--LGDGA--ISALQSYADEDEKESEFYYSWAC-GVNGNPVYAGVKG 108
DB 132 GKDRVSYIECEERSTGQESGIRLQVYADPDPTDESPFCAMGSDVYTGDPVLAAGYRG 191
QY 109 IIVIVDNSTTIKSLVGHDSVNEIRTOPLKPOLYITASKDSVRLMNVETGICLIFA 168
DB 192 VIRIFNPVKHQCSKNYIGHHAINELEKFPHTRPOLLGSKDSRLRMNIQSDVCAVFG 251
QY 169 GAGHRYEVLSDVDFHPSDIY--RFASGMDTYIKIM----SMKEFWYKESFTWDDP 221
DB 252 GVGHGHDVLSVDF---DLNGDRIMSSGMDHSLKLRKLPDIKEA-IELSSGFSPNKNT 307
QY 222 SKRPTFVQFPV--TASIHNTVYDCNMFQDFILSKSVNDELILMEP-OLKEN----SPG 275
DB 308 GPRPTKEHPRDSTDRHNNYDVQVFMDFEFSKSCNSIYCMKRGKLSSEWHETKP 367
QY 276 EGASDVLRIPVPMCDIWFKEGCDLHLSVAIGNQEGKYVWDKSCPYLV--ITKLSH 333
DB 368 ESATVYVLMHFDYKMCETWFRFAFNAMQKILLAGNDLGTTFWELDCNDNLTKCSQLVH 427
QY 334 NQSKSVIROTAMVSDSTILACEDGTIRMPD 365
DB 428 PKNSNITRQTSFKSDGSIIVCVCDDSTWRWD 459

```

```

RESULT 9
Q90YL5 PRELIMINARY; PRT; 426 AA.
AC Q90YL5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE EMBRYONIC ECTODERM DEVELOPMENT PROTEIN (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodidae; Xenopus.
NCBI_TaxId=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21106381; PubMed=11158321;
RA Saitin D.P., Hamer K.M., den Blaauwen J., Otte A.P.;
RT "The polycomb group protein EED interacts with YY1, and both proteins
RT induce neural tissue in Xenopus embryos.";
RL Mol. Cell. Biol. 21:1360-1369(2001).
DR EMBL: AF034136; AAK59991.1; -
FT NON_TER 426
SQ SEQUENCE 426 AA; 48594 MW; 9C862D524E0780F0 CRC64;

```

```

Query Match 36.3%; Score 717; DB 13; Length 426;
Best Local Similarity 41.3%; Pred. No. 1.2e-58;
Matches 151; Conservative 69; Mismatches 120; Indels 26; Gaps 12;
QY 20 KSKYKTYNRIOEG-KRPLAVVNFILARFDF--VFVTAGGNRTIYNCIGDAISALQS 76
DB 64 KYFCKVNSLKEHNPRLRGVGFNW-HSKRGDPLVFATVGSNNVTIYEGHPQGDILNLS 122

```

Oy	77	YADDKESFEYTVSMAGV- GNPYVAAGVKIIRIVDNSTHKSILVGHDSVNER	133
Db	123	YVADADENFTCTMYDTSNTHPLVAASRGRIILINITMOCIKHYGHGNAINEK	182
Oy	136	TOPLPOLVITASKDESVRMLNVEETGICILIFAGAGCHREVLSPVDFHPSDI-- -FPASC	193
Db	183	FHPFDPNILLSVSDHRLRLMNIOTDPLVAFGVGEHREVLISADY--- -DLGKIMSG	239
Oy	194	GMDTTIKIMSK--EFWTVYKSEFTWDDPSK--- -FPFKVOPVF-TASHTNVDCN	246
Db	240	GMDSHLKLMLRINSLRMTAKESYD-- -NENKTRPVSOKVHPFSTRDHIRNVDCV	297
Oy	247	RWGEFITSKSVNDIELMEP----- -QKENSPEGASDVLTAPVPMOIMHFKTSCD	301
Db	298	RWLGLILSKSCENAIYCMWKGKMEDEIKIKASESNVTLIGRDVSCQIMYRSMDF	357
Oy	302	HLSSVAIGNOGKYVMDLKSCP-- -VLITKLSHNSOKSVIRPOTAMSVDSTTIACCEDG	359
Db	358	WOKMALTENOVGKILVMDLEFKDPHKKACTITTYPKCASALRQSFERSVLTAAVODS	417
Oy	360	TWRWD 365	
Db	418	TWRWD 423	
RESULT	10		
O24338			
AC	O24338	PRELIMINARY;	PRT; 425 AA.
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	ESC OR CGE14941.		
GN	ESC OR CGE14941.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:		
OC	Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:		
OC	Ephydroidea: Drosophilidae: Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BERKELEY.		
RA	MEDLINE=20196606; PubMed=10731132;		
RA	Adams M.D., Gelinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	Angelidis P., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Mortman J.-H., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-R., Blazek R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abdill J.F., Abdyaoui A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,		
RA	Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos J., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.		
RA	Durkin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.		
RA	Flores C., Gabrielian A.E., Garg N.G., Gelbart W.M., Glasser K.,		
RA	Glosker A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Honck J.,		
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jostali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., Mcherson J.,		
RA	Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Pacle J.M.,		
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		

[illegible]


```

RESULT 13
ID 016021 PRELIMINARY: PRT: 412 AA.
AC 016021:
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE EXTRA SEX COMBS.
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Papilionoidea; Nymphalidae; Nymphalinae; Junonia.
OX NCBI_TaxID=39708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98001595; PubMed=9343430;
RA Ng J., Li R., Morgan K., Simon J.;
RT "Evolutionary conservation and predicted structure of the Drosophila
RT extra sex combs repressor protein."
RL Mol. Cell. Biol. 17:6663-6672(1997).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF003603; AAC05331.1; -.
DR InterPro: IPR001005; Myb_DNA_bind.
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR PRINTS: PR00320; GPOTEINBRPT.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 412 AA; 46757 MW; 10D47281BDFE2BC6 CRC64;

Query Match 35.0%; Score 690; DB 5; Length 412;
Best Local Similarity 38.0%; Pred. No. 3,7e-56;
Matches 150; Conservative 71; Mismatches 134; Indels 40; Gaps 12;

QY 5 TLGNESTVSTGTPSNK-----SYKVTNRI-----QEGKKPLAAVFN-FLD 45
DB 17 TSNTDNTSRSETPNTTRVKKRRRKKRKKVTKQVKPKFNCSAKEDHGQPLTGCOFNHHLG 76
QY 46 ARFDVFTAGNRTITVNCIGDAISALQSYADEDEKESEFYVSWAGVNGN-PYVAAG 104
DB 77 EGEPLVFAVGSNNKVSVECPESGFKLQCYADPDVDETTYTCAMSYEEETNPLAVALA 136
QY 105 GVKGIIRVADNSETIHKSLVGHGDSVNEIRTPOLKPOLVITASKDESVRILMNVETGICI 164
DB 137 GSRGIIRVFHATQCTCIKHVGHGHAINEVFKHPDPMILLSSAKDHALRLMINISDPCI 196
QY 165 LIFPAGAGHREVLVSDFPHSDI--YRFASGGMOTTIKI-----SMKEFTYVKSFTW 217
DB 197 AIFGVEGEHREVLVSADP---DLGERITMCGMHSKILMRLDKSME--AIKQSYNF 250
QY 218 TDDPSKFTKFPVQ--FPVF--TASIHNTVDCNRMGDFILSKSVNELLMEPOLKENS- 273
DB 251 NPHRALRFPNSIKHEFPFSTRDTHRNVTDCVYRMGDLILSKSCNAILITCKMFGLEDETE 310
QY 274 --PGEAGSDVLLRYFVPMCDIWFIFKSCDLHLSSVAIGNOEGKYYVMDLKS--CPVLLTK 330
DB 311 LRPQGNSTYIYHRPDKCEIWFIFRAVDYSQRIALGNQCKTWMLGSAAGSRVSO 370
QY 331 LSHNOSKSVIRQTASVDSGTTILACCEGDTIWRMD 365
DB 371 LVHPRCAVAARQVTLSTRNGKYLTLCCDDGTIWRMD 405

```

```

ID 016022 PRELIMINARY: PRT: 437 AA.
AC 016022:
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE EXTRA SEX COMBS.
OS Schistocerca americana (American grasshopper).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridoidea; Acrididae; Schistocerca.
OX NCBI_TaxID=7009;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98001595; PubMed=9343430;
RA Ng J., Li R., Morgan K., Simon J.;
RT "Evolutionary conservation and predicted structure of the Drosophila
RT extra sex combs repressor protein."
RL Mol. Cell. Biol. 17:6663-6672(1997).
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF003604; AAC05332.1; -.
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 4.
DR PRINTS: PR00320; GPOTEINBRPT.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 437 AA; 49655 MW; 7F7C227AADB3E59E CRC64;

Query Match 34.9%; Score 689.5; DB 5; Length 437;
Best Local Similarity 38.6%; Pred. No. 4,5e-56;
Matches 151; Conservative 66; Mismatches 131; Indels 43; Gaps 12;

QY 14 SLTPSN-----KSKYKTNRIQEGK-KELAAVFN-FLDARFDV 52
DB 51 SEPTNRRGKRRGRKKIQQKINNKLQYKFSVSKEDHGQPLTGADPFNHLKAGCPILF 110
QY 53 VTAGNRTITVNCIGDAISALQSYADEDEKESEFYVSWAGVNGN-PYVAAGVKGIIIR 111
DB 111 AAVSNKRTYVECPESGSIKILQCYADPDVDENYTCAMSYEEESGRLAVALAGSRGIR 170
QY 112 VIDVNSSETIHKSLVGHGDSVNEIRTPOLKPOLVITASKDESVRILMNVETGICILFAGAG 171
DB 171 IESPATLSCIRHYIGHGHAINEVFKHPDPMILLSSAKDHALRLMINIKTDVCAIFGVE 230
QY 172 GHRREVLVSDFPHSDIY--RFASGGMOTTIKI-----SMKEFTYVKSFTWDDPSKF 224
DB 231 GHRREVLVSADP---DLGERITMCGMHSKILMRLDKSME--AVANSILFNSARSILR 284
QY 225 PTKFVQ--FPVF--TASIHNTVDCNRMGDFILSKSVNELLMEP--QIKENSPEGGA 278
DB 285 PPDLSKEHFPFSTRDTHRNVTDCVYRMGDLILSKSCNAILITCKMFGLEKELKTNDTN 344
QY 279 SDVLLRYFVPMCDIWFIFKSCDLHLSSVAIGNOEGKYYVMDLKSPP--VLTILSHNOS 336
DB 345 VTIHREYRCEIWFVFAFADFQKILALGNQCKTFVMDLVSDPSQSRCTALTTHPRC 404
QY 337 KSVIRQTASVDSGTTILACCEGDTIWRMDVI 367
DB 405 VAAIRQTSLSRDGSVLLCCDDGTIWRMDRI 435

```

RESULT 14
016022

RESULT 15
Q95SW6 PRELIMINARY: PRT: 190 AA.
AC Q95SW6:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE SD02661P.
GN CG5202.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.,
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY060456; AAL25495.1; -
SQ SEQUENCE 190 AA; 22015 MW; 5EBE1A3B99D06E14 CRC64;

Query Match 18.8%; Score 370; DB 5; Length 190;

Best Local Similarity 41.2%; Pred. No. 1.2e-26;

Matches 77; Conservative 33; Mismatches 63; Indels 14; Gaps 6;

QY 192 SCGMDFTIRI-----SMKEFWTVVEKSFMTDDPSKFPKFOFPVF-TASIHFNVDG 245
DB 2 SSGMDHSLKMRIDKPDKEA-IELSSGFSFNKNTGPFPIKEHFDSTROIHRNVDG 60
QY 246 NRWFGDFILSKSVNEILMEP-QLKEN---SPGEGASDVLLRYPPVPCDIWFIKFS 300
DB 61 VWFVGDFEVSFSCSNCISIVCWPGLSESWHEIKPQESATVLIHFDYKMCIEIWFVRAFN 120
QY 301 LHSVAIGNOEKQVYVMDLSCPYL--ITKISHNOSKSVIRQTAMSVDSGTTIACED 358
DB 121 AMOKIILALGNQIGTFVWELDCNDPNLTKCSQLVHPKSNSTIROTSFSKDGSIIVCVCDD 180
QY 359 GTIWRMD 365
DB 181 STIWRMD 187

Search completed: June 25, 2002, 01:24:47
Job time: 651 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 00:14:26 ; Search time 67.93 Seconds

(without alignments)
132.681 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973

Sequence: 1 MSKITLGNSIVSLRPSNK.....STLLACEDGTIRMDVITK 369

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1973	100.0	369	4	US-09-177-249-4
2	202.5	10.3	704	1	US-08-190-802A-62
3	202.5	10.3	704	1	US-08-188-582-5
4	202.5	10.3	704	1	US-08-646-715-5
5	202.5	10.3	704	2	US-08-308-818-3
6	202.5	10.3	704	4	US-08-477-346-62
7	202.5	10.3	704	4	US-08-473-089-62
8	188	9.5	798	1	US-08-190-802A-64
9	188	9.5	798	1	US-08-190-802A-68
10	188	9.5	798	2	US-08-308-818-2
11	188	9.5	798	4	US-08-477-346-64
12	188	9.5	798	4	US-08-477-346-68
13	188	9.5	798	4	US-08-473-089-64
14	188	9.5	798	4	US-08-473-089-68
15	176.5	8.9	517	4	US-08-190-802A-30
16	176.5	8.9	517	4	US-08-477-346-30
17	176.5	8.9	517	4	US-08-473-089-30
18	175.5	8.9	1146	4	US-08-914-999-6
19	170.5	8.6	514	4	US-08-190-802A-66
20	170.5	8.6	514	4	US-08-477-346-66
21	170.5	8.6	514	4	US-08-473-089-66
22	168	8.5	409	2	US-08-283-917-3
23	168	8.5	409	2	US-08-961-716-3
24	168	8.5	410	2	US-08-283-917-9
25	168	8.5	410	2	US-08-961-716-9
26	167.5	8.5	704	1	US-08-188-582-18
27	167.5	8.5	704	1	US-08-646-715-18

28	165.5	8.4	409	1	US-08-190-802A-51	Sequence 51, Appl
29	165.5	8.4	409	4	US-08-477-346-51	Sequence 51, Appl
30	165.5	8.4	409	4	US-08-473-089-51	Sequence 51, Appl
31	163	8.3	713	1	US-08-190-802A-63	Sequence 63, Appl
32	163	8.3	713	4	US-08-477-346-63	Sequence 63, Appl
33	163	8.3	713	4	US-08-473-089-63	Sequence 63, Appl
34	162.5	8.2	906	1	US-08-190-802A-31	Sequence 31, Appl
35	162.5	8.2	906	4	US-08-477-346-31	Sequence 31, Appl
36	162.5	8.2	906	4	US-08-473-089-31	Sequence 31, Appl
37	157.5	8.0	375	4	US-09-063-743-1	Sequence 31, Appl
38	153.5	7.8	317	1	US-08-190-802A-27	Sequence 27, Appl
39	153.5	7.8	317	1	US-08-190-802A-41	Sequence 41, Appl
40	153.5	7.8	317	1	US-08-190-802A-47	Sequence 47, Appl
41	153.5	7.8	317	4	US-08-477-346-27	Sequence 27, Appl
42	153.5	7.8	317	4	US-08-477-346-41	Sequence 41, Appl
43	153.5	7.8	317	4	US-08-473-089-47	Sequence 47, Appl
44	153.5	7.8	317	4	US-08-473-089-27	Sequence 27, Appl
45	153.5	7.8	317	4	US-08-473-089-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1									
US-09-177-249-4									
Sequence 4, Application US/09177249									
Patent No. 6229064									
GENERAL INFORMATION:									
APPLICANT: Fischer, Robert L.									
APPLICANT: Ohad, Nir									
APPLICANT: Kiyosue, Tomohiro									
APPLICANT: Yadegar, Ramon									
APPLICANT: Margossian, Linda									
APPLICANT: Harada, John									
APPLICANT: Goldberg, Robert B.									
TITLE OF INVENTION: The Regents of the University of California									
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit									
FILE REFERENCE: 023070-08612005									
CURRENT FILING DATE: 1998-10-22									
EARLIER APPLICATION NUMBER: US 09/071, 838									
EARLIER FILING DATE: 1998-05-01									
NUMBER OF SEQ ID NOS: 324									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 4									
LENGTH: 369									
TYPE: PR									
ORGANISM: Arabidopsis sp.									
US-09-177-249-4									
Query Match									
Best Local Similarity 100.0%; Score 1973; DB 4; Length 369;									
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MSKITLGNSIVSLRPSNKSKSYKVTNRIOEGKKPLVAVFNLARFDFVYTAGNRI	60						
DB	1	MSKITLGNSIVSLRPSNKSKSYKVTNRIOEGKKPLVAVFNLARFDFVYTAGNRI	60						
QY	61	TYLNCGLDGLAISLQSYADDEKESFYYSVACVGNVYVAAAGVKKGIIRYDVNSPT	120						
DB	61	TYLNCGLDGLAISLQSYADDEKESFYYSVACVGNVYVAAAGVKKGIIRYDVNSPT	120						
QY	121	HKSLVGHGDSVNEIRPQPLKPOLVITASKDESRYRLNVEFGICILIFAGAGHRYEVL	180						
DB	121	HKSLVGHGDSVNEIRPQPLKPOLVITASKDESRYRLNVEFGICILIFAGAGHRYEVL	180						
QY	181	DEHPSDIYFPAQSGMDTITIKSMKEFWTVYKESFTWDDPSKPKFYQVFTASHT	240						
DB	181	DEHPSDIYFPAQSGMDTITIKSMKEFWTVYKESFTWDDPSKPKFYQVFTASHT	240						
QY	241	NYVDCNRMGDFILSKSVNENILMERPOLKENSFGGASADVLIRYVPRCDITFI	300						

Db	241	NYVDCNRFVGGFLLSKSYDNEIILMEPQLKENSPEGASDVLLRLPYVMCDIMEIFKSCD	3000
Qy	301	LHLSSVAIGNEGKYVWDLKSCPEVLTIKLSHQSSVIRQTAMSVDSSTIIACCEGDT	3600
Db	301	LHLSSVAIGNEGKYVWDLKSCPEVLTIKLSHQSSSVIRQTAMSVDSSTIIACCEGDT	3600
Qy	361	IWRMDVITK	369
Db	361	IWRMDVITK	369

RESULT 2

```

US-08-190-802A-62
Sequence 62, Application US/08190802A
Patent No. 5519003
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Delinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A
FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8600-0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO. 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 704 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANSCRIPTION FACTOR TTF, Fig. 43
US-08-190-802A-62

```

Query Match	10.3%	Score 202.5	DB 1	Length 704
Best Local Similarity	25.2%	Pred No. 2.2e-13		
Matches 65; Conservative	49	Mismatches 99	Indels 45	Gaps 13

```

OY      110 IVIDVNSETIHKSLVYGDDSVNEIRTOPLEPOLVITAKSKEVSLWNNVEGICILIFAG 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      433 YMLDDRSGEYTRSRIMSGHTGEPYRCFAFP-EMNILLSCSEDSSTRILMSLLTWSCVYTR- 489
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      170 AGCHHYEVLSDVDEHSDIYLRASGDMOTIKIWSKKEWTVYKSLFWTDDPSKFPFKV 229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      490 --GHYYPWMDRFAFPHGYT-FVSCSYDTAKIYMA-----TDSNQ----- 525
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      230 QEPVFTASIHNTYDCNRF--GDEILSKSVDNELLMEPOLKNSPQEGASDYLLRIPV 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      526 ALRVFVG--HLSDSDVDFHNSPVNVAVAGSSDDRVYRLM--DNMGQGSYVR-LMTGHKG 577
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 288 PMCDIWFIFKFCSDHLHLSVALGNDEGRVYWDLKSCEPVLITKSHQSGSVLROFMASV 347
 Db 578 SVSLSIAF--SNCGRIYAS---GSVDNHNIIIMDL-SNGLSVTLTLRLRH---STVTLTFESR 628
 QY 348 DGSITLACCEGDTIWRMD 365
 Db 629 DGTVLAAGLNDNNTLMD 646

RESULT 3

US-08-188-582-5
Sequence 5, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Rupert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS',
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/ATJ/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-5

Query Match	10.3%	Score 202.5	DB 1	Length 704
Best Local Similarity	25.2%	Pred. No. 2.2e-13		
Matches 65; Conservative	49;	Mismatches 99;	Indels 45;	Gaps 13;

[illegible]

Db 526 ALRFVGG--HLSVDVDCVQFHPNSNYVATGSSDRTVRLM-----DNMTGQSVR-LMTGCHKG 577
QY 288 PMCDIWFIFKSCDLHLSVAIGNQEGKYVWDLKSCPVLITKLSHNSQSVIRQTAMSV 347
Db 578 SVSSLAFF--SACGRYLAS--GSVDHNIITWDL-SNGSLVTTILRHT---STVTTTTFFSR 628
QY 348 DGSSTILACCEGDTIWRMD 365
Db 629 DGTVLAAAGLDNNLTLM 646

RESULT 4

US-08-646-715-5
; Sequence 5, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dylact, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-646-715-5

Query Match 10.3%; Score 202.5; DB 1; Length 704;
Best Local Similarity 25.2%; Pred. No. 2.2e-13;
Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;
QY 110 IRVIDVNSFIHKSILVGHGDSVNEIRTPQLKPOLVITASDESVRIMNVETGICILIFAG 169
Db 432 VRRLDRSGEVTRSLMGHGDPYRCAPAP-EMNLLSCSDSTIRLMSLTITWSCVVTYR- 489
QY 170 AGGHRREVLSVDHPSPDIYFAFASCGMDTTIKINSMEFWTVYVKSFTWTDPSKFPYK 229

Db 490 --GHVYFVMDVREAPRGY-FVSCSYDKTARLMA-----TDSNQ----- 525
QY 230 QFVVFPASTHTNVDDCNRMF--GDFILSKVDNDEILMEPQLKENSPEASVLLRYPV 287
Db 526 ALRFVGG--HLSVDVDCVQFHPNSNYVATGSSDRTVRLM-----DNMTGQSVR-LMTGCHKG 577
QY 288 PMCDIWFIFKSCDLHLSVAIGNQEGKYVWDLKSCPVLITKLSHNSQSVIRQTAMSV 347
Db 578 SVSSLAFF--SACGRYLAS--GSVDHNIITWDL-SNGSLVTTILRHT---STVTTTTFFSR 628
QY 348 DGSSTILACCEGDTIWRMD 365
Db 629 DGTVLAAAGLDNNLTLM 646

RESULT 5

US-08-308-818-3
; Sequence 3, Application US/08308818
; Patent No. 5847077
; GENERAL INFORMATION:
; APPLICANT: Green, Michael R.
; APPLICANT: Reese, Joseph C.
; TITLE OF INVENTION: A No. 5847077el Fungal Multisubunit Protein
; TITLE OF INVENTION: Complex critical for Expression of Fungal Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,818
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0342/0A404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-52707700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: D. melanogaster
; IMMEDIATE SOURCE:
; CLONE: TAFIT-80
; US-08-308-818-3

Query Match 10.3%; Score 202.5; DB 2; Length 704;
Best Local Similarity 25.2%; Pred. No. 2.2e-13;
Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;
QY 110 IRVIDVNSFIHKSILVGHGDSVNEIRTPQLKPOLVITASDESVRIMNVETGICILIFAG 169
Db 432 VRRLDRSGEVTRSLMGHGDPYRCAPAP-EMNLLSCSDSTIRLMSLTITWSCVVTYR- 489
QY 170 AGGHRREVLSVDHPSPDIYFAFASCGMDTTIKINSMEFWTVYVKSFTWTDPSKFPYK 229

Db 490 --GHVPVMDVRFAPRGY-FVSCSYDKTARLMA-----IDSNO----- 525
QY 230 QPVPFASITHTNVDCNRF--GDFILSKSVNDEIILMEPOLKENSPEGASVILRYPV 287
Db 526 ALRFVFG--HLSVDCVQFHPNSNYATGSSDRTVRLM-----DNMTGQSVR-IMTGHKG 577
QY 288 PMCDIMEFKSCDLHLSVAIGNQEGKYVWDKSCPPVILTKLSHNSQSVIRQTAMSV 347
Db 578 SVSLSAF--SACGRYLAS--GSVDHNIITIMDL-SNGSLVTILLRHT---STVTTITFSR 628
QY 348 DGSITLACCEGDTIRWD 365
Db 629 DGTVLAAGLDNNLTLMWD 646

RESULT 6

US-08-477-346-62
; Sequence 62, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived peptides and uses
; TITLE OF INVENTION: Theeof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morlison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 704 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANSCRIPTION FACTOR T1IF, Fig. 45
; US-08-477-346-62

Query Match 10.3%, Score 202.5; DB 4; Length 704;
Best Local Similarity 25.2%, Pred. No. 2.2e-13;
Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;
QY 110 IRVIDVNSETHKSLVYGSDSVNEIRPOLKPOLVITASKDSVRLMNVETGICILIFAG 169
Db 432 VMLDDRSGEYTRSLMGHTGPGYRCAPAP-ENMLLLSCSESDTIRLMSLTWSCVYTR- 489

QY 170 AGGHREYLVSDHPSDIYRFASCGMDTTIKIWSMKEFTYVEKSFPTWDDSPKPTREV 229
Db 490 --GHVPVMDVRFAPRGY-FVSCSYDKTARLMA-----IDSNO----- 525
QY 230 QPVPFASITHTNVDCNRF--GDFILSKSVNDEIILMEPOLKENSPEGASVILRYPV 287
Db 526 ALRFVFG--HLSVDCVQFHPNSNYATGSSDRTVRLM-----DNMTGQSVR-IMTGHKG 577
QY 288 PMCDIMEFKSCDLHLSVAIGNQEGKYVWDKSCPPVILTKLSHNSQSVIRQTAMSV 347
Db 578 SVSLSAF--SACGRYLAS--GSVDHNIITIMDL-SNGSLVTILLRHT---STVTTITFSR 628
QY 348 DGSITLACCEGDTIRWD 365
Db 629 DGTVLAAGLDNNLTLMWD 646

RESULT 7

US-08-473-089-62
; Sequence 62, Application US/08473089
; Patent No. 6342368
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived peptides and uses
; TITLE OF INVENTION: Theeof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morlison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,089
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 704 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANSCRIPTION FACTOR T1IF, Fig. 45
; US-08-473-089-62

Query Match 10.3%, Score 202.5; DB 4; Length 704;
Best Local Similarity 25.2%, Pred. No. 2.2e-13;
Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;
QY 110 IRVIDVNSETHKSLVYGSDSVNEIRPOLKPOLVITASKDSVRLMNVETGICILIFAG 169
Db 432 VMLDDRSGEYTRSLMGHTGPGYRCAPAP-ENMLLLSCSESDTIRLMSLTWSCVYTR- 489
QY 170 AGGHREYLVSDHPSDIYRFASCGMDTTIKIWSMKEFTYVEKSFPTWDDSPKPTREV 229

Db 490 --GHVYVMDVREAPHGY--FVSCSYDKTARLMA-----TDSNQ----- 525
 QY 230 QRPVPLASHITNVDCNRF--GDFILSKSVNEILLMEPOLKENSPEGASVLLRY 287
 Db 526 ALRVEFG--HLSVDVDFPHNSNVATGSSDPTRLM-----DNMTGQSVR-LMTGHKG 577
 QY 288 PMCDIFIKFSCDLHLSVAIGNQEGKYVWDLKSCPPYLITKLSINQSKSVIRQTAMSV 347
 Db 578 SVSSIAF--SACGRYLA--GSVDHNIITLMDL-SNGSLVTLTLRHT---SYVTITTFE 628
 QY 348 DGSITLACCEDGTWRWD 365
 Db 629 DGTVLAAGLDNMLTMD 646

RESULT 8

US-08-190-802A-64
 ; Sequence 64, Application US/08190802A
 ; Patent No. 5519003
 ; GENERAL INFORMATION:
 ; APPLICANT: Mochly-Rosen, Daria
 ; APPLICANT: Ron, Dorit
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dehlinger & Associates
 ; STREET: P.O. Box 60850
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-0850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/190,802A
 ; FILING DATE: 01-FEB-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fabian, Gary R.
 ; REGISTRATION NUMBER: 33,875
 ; REFERENCE/DOCKET NUMBER: 8600-0139
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 324-0880
 ; TELEFAX: (415) 324-0960
 ; INFORMATION FOR SEQ ID NO: 64:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 798 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: TUP1 HOMOLOG, Fig. 47
 ; US-08-190-802A-64

Query Match 9.5%; Score 188; DB 1; Length 798;

Best Local Similarity 20.4%; Pred. No. 1,1e-11;

Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

QY 74 LOSYADEDEKEESFYTVSMAGVNGNPNYVAGYKGIIRIVDVNSETH----- 121
 Db 459 MYTFQNTKMDSCLDSDDCRI-----AAAGFQDSYIKIWSLDGSSLNINNALNNNDK 513
 QY 122 ----KSLVGHGDSVNEIRTPQLKPOLVITASKDESVRLMNVETGICILIFAGAGGHRYE 176
 Db 514 EDPCTKTLVGHSGTVYSTSP--DNKYLLSGSDKTVRLMSMDTHALVSK---GHNHP 569

QY 177 VLSVDFPSPDIYRFASCMDPTIKISMKEFWTYVEKSFRTDPSKFPKFEQFPVETA 236
 Db 570 YMVVSFSPGLGH-FATASHQOTARLMSCHI-----YPLRIF 605
 QY 237 SIHTNVDCNRF--GDFILSKSVNEILLMEPOLKENSPEGASVLLRYVPMCDI-- 292
 Db 606 AGHLNDVDCVSPHNCQCYVETGSSDKTCRMWDV-----STGDSVR-LFGLHTAPVISTAV 659
 QY 293 ----WFIKSCDLHLSVAIGNQEGKYVWDLKSCPPYLITKLSINQSKSVIRQTAMSV 347
 Db 660 CPDGRW-----LSTGSDGIINWDTGTGRRL---KRMNGHGNALYSLYSK 704
 QY 348 DGSITLACCEDGTWRWDV 366
 Db 705 EGNVLISGADHTVRYWDL 723

RESULT 9

US-08-190-802A-68
 ; Sequence 68, Application US/08190802A
 ; Patent No. 5519003
 ; GENERAL INFORMATION:
 ; APPLICANT: Mochly-Rosen, Daria
 ; APPLICANT: Ron, Dorit
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dehlinger & Associates
 ; STREET: P.O. Box 60850
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-0850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/190,802A
 ; FILING DATE: 01-FEB-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fabian, Gary R.
 ; REGISTRATION NUMBER: 33,875
 ; REFERENCE/DOCKET NUMBER: 8600-0139
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 324-0880
 ; TELEFAX: (415) 324-0960
 ; INFORMATION FOR SEQ ID NO: 68:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 798 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: Yrb 1410 yeast, Fig. 51
 ; US-08-190-802A-68

Query Match 9.5%; Score 188; DB 1; Length 798;

Best Local Similarity 20.4%; Pred. No. 1,1e-11;

Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

QY 74 LOSYADEDEKEESFYTVSMAGVNGNPNYVAGYKGIIRIVDVNSETH----- 121
 Db 459 MYTFQNTKMDSCLDSDDCRI-----AAAGFQDSYIKIWSLDGSSLNINNALNNNDK 513
 QY 122 ----KSLVGHGDSVNEIRTPQLKPOLVITASKDESVRLMNVETGICILIFAGAGGHRYE 176

Db 514 EDPCTCTLVGSGTGYSTSP-DNKYLLSGSDKTVRLMSMDTHTALVSYK---GHNHP 569
QY 177 VLSVDHPSPDIYFPASGMDTFTIKIMSKEMFTYVEKSFMTDDPSKFTKFPVQFPVETA 236
Db 570 VMDVSPSPGLGHY-FATASHDQTARLMSCDHI-----YPLRIF 605
QY 237 SIHTNYVDCNRWF--GDFILSKSVNDEIILMEPOLKENSFGASDYLRLYPVPMCDI-- 292
Db 606 AGHLNDVDCVSPHNGCYVFTGSSDKTCRMVDY----STGDSVR-LFLGHTAPVISTIAV 659
QY 293 -----WFIKSCDLHLSVAIGNQEGKVYVWDLKSCPPVILITLSHNSQSVIRQTAMSV 347
Db 660 CPDGRW-----LSTGSEDDIINWMDIGTGKRL--KOMRGHKNALYSLSYSK 704
QY 348 DGTIILACEDGTIWRMDV 366
Db 705 EGNVLISGADHTYRVWDL 723

RESULT 10
US-08-308-818-2
; Sequence 2, Application US/08308818
; Patent No. 5847077
; GENERAL INFORMATION:
; APPLICANT: Green, Michael R
; APPLICANT: Reese, Joseph C
; TITLE OF INVENTION: A No. 5847077el Fungal Multisubunit Protein
; NUMBER OF INVENTION: Complex Critical for Expression of Fungal Proteins
; CORRESPONDENCE ADDRESS: 11
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,818
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0342/0A404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-52707700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: S. cerevisiae
; IMMEDIATE SOURCE:
; CLONE: TAF-90
; US-08-308-818-2

Query Match 9.5%; Score 188; DB 2; Length 798;
Best Local Similarity 20.4%; Pred. No. 1,1e-11;
Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;
QY 74 LQSYADEDESESYTVSMAGVNGNPYVAGVGKGIINVIDVNSETH----- 121

Db 459 MYFQNTNRKDMSCLDPSDDCRIT-----AAAGFQDSYIKIKMSLDGSSLNPNNTALNNNDK 513
QY 122 -----KSLVGHGDSVNEIRTPOLPLVITASKDESVRILMNETGICILIFAGAGHRE 176
Db 514 EDPCTCTLVGSGTGYSTSP-DNKYLLSGSDKTVRLMSMDTHTALVSYK---GHNHP 569
QY 177 VLSVDHPSPDIYFPASGMDTFTIKIMSKEMFTYVEKSFMTDDPSKFTKFPVQFPVETA 236
Db 570 VMDVSPSPGLGHY-FATASHDQTARLMSCDHI-----YPLRIF 605
QY 237 SIHTNYVDCNRWF--GDFILSKSVNDEIILMEPOLKENSFGASDYLRLYPVPMCDI-- 292
Db 606 AGHLNDVDCVSPHNGCYVFTGSSDKTCRMVDY----STGDSVR-LFLGHTAPVISTIAV 659
QY 293 -----WFIKSCDLHLSVAIGNQEGKVYVWDLKSCPPVILITLSHNSQSVIRQTAMSV 347
Db 660 CPDGRW-----LSTGSEDDIINWMDIGTGKRL--KOMRGHKNALYSLSYSK 704
QY 348 DGTIILACEDGTIWRMDV 366
Db 705 EGNVLISGADHTYRVWDL 723

RESULT 11
US-08-477-346-64
; Sequence 64, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived peptides and Uses
; NUMBER OF INVENTION: Thereof
; CORRESPONDENCE ADDRESS: 265
; ADDRESSEE: Morison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUP1 HOMOLOG, Fig. 47
; US-08-477-346-64

ORIGINAL SOURCE: yrb 1410 yeast, Fig. 51
INDIVIDUAL ISOLATE:
US-08-477-346-68

```

Query Match          9.5%; Score 188; DB 4; Length 798;
Best Local Similarity 20.4%; Pred. No. 1,1e-11;
Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

QY 74 LQSYADEKKESEFTVSMACGVNGNPYYAAGVKGIIIVIDYNSETH----- 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 MYTFQNTKMKKSCUDFSDCRI-----AAAFQDSYIKWISLDGSSLNPINALNNDKD 513
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 -----KSLVGHGDSVNEIRTOPLKPOLVITFASKDESVRLMNVETICILIFAGAGHRYE 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 514 EDPPICKTILVGHSGVYISTFSFP-DNKYLLSGSEDKTVLMSMDHTALVSKR--GHNHP 569
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 VLSYDFHPSDILYRFASCGMDFTTIKWSKMEFTWYVEKSFPTWDDPSKEPTFVDFPEVETA 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 570 VMVDVSFSPLGHY-FATASHDQTARLMSCDHI-----YPLNLF 605
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 237 SIHTNYVDCNRMF--GDFILSKSVDNELLWEPEOLKENSPEGASVDLLRVPVPCDI-- 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 AGHLNDVDCVAFHPRNGCVYFTGSSDKTCRMDV-----STGDSVR-LFLGHTAPYISIAV 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 -----WPIKFGSCDLHLSSVALIGNQDEGKYYVMDLAKSCPVLVLTIKLSHNSKSVIRQTASV 347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY      348  DSGTILACCEGDTIRWMD 366
      :|:::| |::| |
Db      705  EGNVLISGCAHDTYRWMDL 723

RESULT 13
US-08-473-089-64
: Sequence 64, Application US/08473089
: Patent No. 6342368
: GENERAL INFORMATION:
: APPLICANT: Mochly-Rosen, Daria
: APPLICANT: Ron, Dorit
: TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
: TITLE OF INVENTION: Thereof
: NUMBER OF SEQUENCES: 265
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Morrison & Foerster
: STREET: 2000 Pennsylvania Avenue, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20006-1812
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/473,089
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURASHIGE, KATE H.
: REGISTRATION NUMBER: 29,959
: REFERENCE/DOCKET NUMBER: 2550-0025.22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763
: INFORMATION FOR SEQ ID NO: 64:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 798 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein

```

TYPE: amino acids

```

; INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:

```

```

; INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 00:11:36 ; Search time 79.11 Seconds
(without alignments)
518.092 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973

Sequence: 1 MSKITLGNSIVSLRPSNK.....STILACCEGDTIMRWVITK 369

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_032802:*

1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1973	100.0	369	21	AAV57037
2	1586	80.4	372	22	AAU00313
3	1576	79.9	370	22	AAU00315
4	1501	76.1	387	22	AAU00326
5	1402.5	71.1	379	22	AAU00309
6	1402.5	71.1	379	22	AAU00323
7	1394.5	70.7	379	22	AAU00321
8	1394	70.7	378	22	AAU00327
9	1391	70.5	375	22	AAU00325
10	1337.5	67.8	391	22	AAU00317
11	1219.5	61.8	461	22	AAU00311

12	1219.5	61.8	461	22	AAU00322	Fertilisation-inde
13	1215.5	61.6	363	22	AAU00339	Fertilisation-inde
14	1103	55.9	273	22	AAU00318	Fertilisation-inde
15	1045	53.0	430	22	AAU00320	Fertilisation-inde
16	720	36.5	162	22	AAU00342	Fertilisation-inde
17	718	36.4	462	22	ABB61196	Drosophila melanog
18	715	36.2	425	22	ABB69961	Drosophila melanog
19	708	35.9	158	22	AAU00335	Fertilisation-inde
20	662	33.6	152	22	AAU00336	Fertilisation-inde
21	648	32.8	141	22	AAU00341	Fertilisation-inde
22	590	29.9	145	22	AAU00330	Fertilisation-inde
23	590	29.9	155	22	AAU00312	Fertilisation-inde
24	507	25.7	129	22	AAU00334	Fertilisation-inde
25	498	25.2	125	22	AAU00314	Fertilisation-inde
26	486.5	24.7	119	22	AAU00332	Fertilisation-inde
27	462	23.4	121	22	AAU00316	Fertilisation-inde
28	456	23.1	124	22	AAU00333	Fertilisation-inde
29	422	21.4	119	22	AAU00324	Fertilisation-inde
30	365	18.5	98	22	AAU00340	Fertilisation-inde
31	335.5	17.0	116	22	AAU00329	Fertilisation-inde
32	309	15.7	111	22	AAU00343	Fertilisation-inde
33	264.5	13.4	84	22	AAU00328	Fertilisation-inde
34	220	11.2	59	22	AAU00331	Fertilisation-inde
35	216	10.9	59	22	AAU00338	Fertilisation-inde
36	202.5	10.3	704	16	AA85877	WP-40 domain-contg
37	202.5	10.3	704	17	AA85879	Drosophila TATA-bi
38	202.5	10.3	704	18	AA85877	TATA-binding prote
39	202.5	10.3	704	22	ABB63447	Drosophila melanog
40	198	10.0	55	22	AAU00337	Fertilisation-inde
41	198	10.0	55	22	AAU00319	Fertilisation-inde
42	197.5	10.0	704	15	AA856488	TATA-binding prote
43	194.5	9.9	542	21	AA856488	Human E3 ubiquitin
44	194.5	9.9	542	22	AA856488	Human protein SEQ
45	194.5	9.9	542	22	AA856488	Human polypeptide

ALIGNMENTS

RESULT 1	AAV57037	standard: Protein; 369 AA.
ID	AAV57037	
XX	AAV57037;	
AC	21-FEB-2000 (first entry)	
XX		
DT	Fertilisation-independent endosperm 3 (FIE3) amino acid sequence.	
XX		
DE	FIE3; fertilisation-independent endosperm; plant regulatory gene;	
XX	reproductive development; polycomb gene; transgenic plant;	
KW	plant morphology; cell cycle control; increase gene expression;	
KW	vegetative growth.	
XX		
OS	Arabidopsis sp.	
XX		
PN	W09957247-A1.	
XX		
PD	11-NOV-1999.	
XX		
PF	03-MAY-1999; 99MO-US09676.	
XX		
PR	01-MAY-1998; 98US-0071838.	
XX	22-OCT-1998; 98US-0177249.	
PA	(RECC) UNTV CALIFORNIA.	
XX		
PI	Fischer RL, Ohad N, Kiyosue T, Yadegari R, Margossian L, Harada J;	
XX	Goldberg RB;	
DR	WPI; 2000-023577/02.	
DR	N-PSDB; AA237896.	
XX		

[illegible][illegible]

Db 16 lgcgvvyslacsakreyivtrnlqegkrplyavlnflisryfnvatvgmrvtlygc 75
 QY 66 LGGCAISALOSYADEDEKESEFTYVSNACGVNPNPYAAGVKGIIRYIDVNSERTIHKSLV 125
 Db 76 ldegdlavlsyadedeknesfylvgaenvgdrcplvvaaglnvlyiridagsekthsfv 135
 QY 126 GHGDSVNEIRTOPLKPOLVITASKDESVRLMNVETGICILIFAGAGHREYLSVDFHP 185
 Db 136 ghgdsinevkaqlnpllvvasakdesirlnwhtgicilifagagghrnevlsvdfhps 195
 QY 186 DIYRFASCGMDTITIKIMSKKEFTYVSKFTWDDPSKPTKVFQPFVFTASIHNTVDC 245
 Db 196 dmyricscgmndstvkfwsmkeftwlyekstwtclpskfpkfyqfvtasvhnlyvdc 255
 QY 246 NRWFGEFTLSKSVNELLMEPOLKENSPEGASDVLRLYVPMCDIMETIFSCDHLSS 305
 Db 256 nrwlgdflilsksvdneilwepkneqtpkyvvdvlhkyipidcnlwlkfsdchfn 315
 QY 306 VAIGNQEGKYVWDLKSCPPVLITKLKSHNOSKSVIRQTAMSYDGSITLACCEDGTIWRMD 365
 Db 316 vtvguregkllfwelqgsppvllaaklshpskspirqtatsfdgsitlsccegdgtiwrwd 375
 QY 366 V 366
 Db 376 V 376
 RESULT 5
 AAU00309 standard; Protein: 379 AA.
 AC AAU00309;
 DT 12-SEP-2001 (first entry)
 Db Fertilisation-independent endosperm protein, ccase-b.pk0026.g4.
 XX Fertilisation-independent endosperm; plant reproduction; apomixis;
 KM seed; pharmaceutical; nutraceutical; polymer; ccase-b.pk0026.g4; corn.
 OS Zea mays.
 XX WO200116325-A2.
 PN 08-MAR-2001.
 PD 30-AUG-2000; 2000WO-US23735.
 PF 31-AUG-1999; 99US-0151575.
 PR (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;
 PI Simmons CR, Weng Z;
 XX WPI: 2001-244407/25.
 DR N-PSDB: AAS01156.
 PT New plant fertilization independent endosperm protein for the
 PT production of seed without fertilization is recombinantly produced -
 XX Claim 1; Page 39-40; 94pp; English.
 PS The sequence represents the amino acid sequence of fertilisation-
 XX independent endosperm protein, encoded by cDNA clone ccase-b.pk0026.g4,
 CC isolated from Zea mays. Fertilisation-independent endosperm proteins are
 CC plant reproduction proteins necessary for apomixis, the formation of
 CC seeds without fertilisation. Apomixis is especially useful to agriculture
 CC because it eliminates the necessity of selfing plants to produce
 CC genetically identical seed. Such seed is useful to produce seeds for
 CC human and animal food and for commercial milling and extraction,
 CC including the production of useful recombinant products in the endosperm

CC e.g.pharmaceutical, nutraceutical, industrial compounds and polymers.
 CC Embryoless seed production by transgenic plants is less likely to cause
 CC ethical and environmental concern over transgenic plant production as no
 CC gametes are being produced to cross pollinate with other crops and the
 CC seeds cannot germinate.
 XX Sequence 379 AA:
 SQ
 Query Match 71.1%; Score 1402.5; DB 22; Length 379;
 Best Local Similarity 69.4%; Pred. No. 5.5e-147;
 Matches 252; Conservative 48; Mismatches 62; Indels 1; Gaps 1;
 QY 6 LGNESIVSLTPSNKSKSYVTNRIOEGKRPlyAVENFELDARFDFVFTAGNRITLYNC 65
 Db 10 lgcnaaegslvpsrkykpcgkhtegkrplyaigfnmdarydvfatvgmrvtlyrc 69
 QY 66 LGGCAISALOSYADEDEKESEFTYVSNACG-VNCPYPAAGVGIIRYIDVNSERTIHKSL 124
 Db 70 lengsfallqayvedekdesfytlswardhvdspilvaagsngilrvlnccateklaksf 129
 QY 125 VGHGDSVNEIRTOPLKPOLVITASKDESVRLMNVETGICILIFAGAGHREYLSVDFHP 184
 Db 130 vghgdsineirtpkpsllisaskdesvrlmwhtgicilifagagghrnevlsvdfhps 189
 QY 185 SDIYRFASCGMDTITIKIMSKKEFTYVSKFTWDDPSKPTKVFQPFVFTASIHNTVDC 244
 Db 190 sdierfascgmdntvkfwsmkeftwlyekstwtclpskfpkfyqfvtasvhnlyvdc 249
 QY 245 CNRWFGEFTLSKSVNELLMEPOLKENSPEGASDVLRLYVPMCDIMETIFSCDHLSS 304
 Db 250 ctnrwlgdflilsksvdneilwepkneqtpkyvvdvlhkyipidcnlwlkfsdchfn 309
 QY 305 SVAIGNQEGKYVWDLKSCPPVLITKLKSHNOSKSVIRQTAMSYDGSITLACCEDGTIWRMD 364
 Db 310 qlaigntregklywewegsppvllarllyngqckspirqtatsfdgsitlsccegdgtiwrwd 369
 QY 365 DVI 367
 Db 370 dev 372
 RESULT 6
 AAU00323 standard; Protein: 379 AA.
 ID AAU00323
 AC AAU00323;
 DT 12-SEP-2001 (first entry)
 Db Fertilisation-independent endosperm protein, p0107.cbca179r.
 XX Fertilisation-independent endosperm; plant reproduction; apomixis;
 KM seed; pharmaceutical; nutraceutical; polymer; p0107.cbca179r.
 OS Zea mays.
 XX WO200116325-A2.
 PN 08-MAR-2001.
 PD 30-AUG-2000; 2000WO-US23735.
 PF 31-AUG-1999; 99US-0151575.
 PR (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;
 PI Simmons CR, Weng Z;
 XX WPI: 2001-244407/25.
 DR N-PSDB: AAS01170.

XX New plant fertilization independent endosperm protein for the
PT production of seed without fertilization is recombinantly produced -
PS Claim 1; Page 63-64; 94pp; English.

XX The sequence represents the amino acid sequence of fertilisation-
CC independent endosperm protein, encoded by cDNA clone p0107.cbca479r.
CC Fertilisation-independent endosperm proteins are plant reproduction
CC proteins necessary for apomixis, the formation of seeds without
CC fertilisation. Apomixis is especially useful to agriculture because it
CC eliminates the necessity of selfing plants to produce genetically
CC identical seed. Such seed is useful to produce seeds for human and animal
CC food and for commercial milling and extraction, including the production
CC of useful recombinant products in the endosperm e.g. pharmaceutical,
CC nutraceutical, industrial compounds and polymers. Embryolless seed
CC production by transgenic plants is less likely to cause ethical and
CC environmental concern over transgenic plant production as no gametes are
CC being produced to cross pollinate with other crops and the seeds cannot
CC germinate.

XX Sequence 379 AA:

Query Match 71.1%; Score 1402.5; DB 22; Length 379;
Best Local Similarity 69.4%; Pred. No. 5-5e-147;
Matches 252; Conservative 48; Mismatches 62; Indels 1; Gaps 1;

OY 6 LGNESIVGSLTPSNKSKYKVTNRIOEGKKPLVAVVFNFLDARFEDVFTAGNRLTYNC 65
DB 10 lgcgaagslvparkreykpcgkhegkrlpalyaigfnmdarydvfatvggnrvctlyc 69
OY 66 LGDGAISALQSYADDEKESFYTVSWACG-VNGNPNVAGVKGITRVIDVNSERTHKSL 124
DB 70 lengsfallqayvdedkdesfylvswardhvdgspllvaaagngilrvincateklaksf 129
OY 125 VGHGDSVNEIRPOLPLPOLVITASKDESVRLMNVETGICILIFAGAGHREYLSVDFHP 184
DB 130 vghgdsineirtpkplksllsaskdesvrlmnhvgicillifagaghnreylsvdfhp 189
OY 185 SDIYRFASCGMDTTIKMSKKEFWTVYVEKSFMTWDDPSKPFKEVQFPVFTASIHNTYVD 244
DB 190 sdierfascgmdntvkiwskewfwvtdpskpfkcygfpvllaaahsnnyd 249
OY 245 CNRMFGDFTLSKSVNELLMEPOLKENSPEGASDVLLRYVPMKDIWIFRSCDLHLS 304
DB 250 ctnwlgdftllsksvndelwepkteqspgsidllqkyppecdiwifkscdfhn 309
OY 305 SVAIGNOEGKVYVMDLSCPPVILITKLSHNSQSVYRQTAMSVDSSTILACCEDGTIRMW 364
DB 310 qiaignregklywewgspvllarlynqgckspirqlavsfdsstllgagedgtlwrw 369
OY 365 DVI 367
DB 370 dev 372

RESULT 7

AAU00321 AUU00321 standard; Protein; 379 AA.

AC AAU00321;

DT 12-SEP-2001 (first entry)

DE Fertilisation-independent endosperm protein, p0101.cgamg48r.

KM Fertilisation-independent endosperm; plant reproduction; apomixis;

XX seed; pharmaceutical; nutraceutical; polymer; p0101.cgamg48r.

OS Zea mays.

PN WO200116325-A2.

XX 08-MAR-2001.

XX 30-AUG-2000; 2000MO-US23735.

XX 31-AUG-1999; 99US-0151575.

PA (DUPO) DU PONT DE NEMOURS & CO E. I.
PA (PION-) PIONEER HI-BRED INT INC.

PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
PI Simmons CR, Weng Z;

DR WPI: 2001-244407/25.

DR N-PSDB: AAS01168.

PT New plant fertilization independent endosperm protein for the
PT production of seed without fertilization is recombinantly produced -
PS Claim 1; Page 58-60; 94pp; English.

CC The sequence represents the amino acid sequence of fertilisation-
CC independent endosperm protein, encoded by cDNA clone p0101.cgamg48r.
CC Fertilisation-independent endosperm proteins are plant reproduction
CC proteins necessary for apomixis, the formation of seeds without
CC fertilisation. Apomixis is especially useful to agriculture because it
CC eliminates the necessity of selfing plants to produce genetically
CC identical seed. Such seed is useful to produce seeds for human and animal
CC food and for commercial milling and extraction, including the production
CC of useful recombinant products in the endosperm e.g. pharmaceutical,
CC nutraceutical, industrial compounds and polymers. Embryolless seed
CC production by transgenic plants is less likely to cause ethical and
CC environmental concern over transgenic plant production as no gametes are
CC being produced to cross pollinate with other crops and the seeds cannot
CC germinate.

SO Sequence 379 AA:

Query Match 70.7%; Score 1394.5; DB 22; Length 379;
Best Local Similarity 68.9%; Pred. No. 4.3e-146;
Matches 250; Conservative 48; Mismatches 64; Indels 1; Gaps 1;

OY 6 LGNESIVGSLTPSNKSKYKVTNRIOEGKKPLVAVVFNFLDARFEDVFTAGNRLTYNC 65
DB 10 lgcgaagslvparkreykpcgkhegkrlpalyaigfnmdarydvfatvggnrvctlyc 69
OY 66 LGDGAISALQSYADDEKESFYTVSWACG-VNGNPNVAGVKGITRVIDVNSERTHKSL 124
DB 70 lengsfallqayvdedkdesfylvswardhvdgspllvaaagngilrvincateklaksf 129
OY 125 VGHGDSVNEIRPOLPLPOLVITASKDESVRLMNVETGICILIFAGAGHREYLSVDFHP 184
DB 130 vghgdsineirtpkplksllsaskdesvrlmnhvgicillifagaghnreylsvdfhp 189
OY 185 SDIYRFASCGMDTTIKMSKKEFWTVYVEKSFMTWDDPSKPFKEVQFPVFTASIHNTYVD 244
DB 190 sdierfascgmdntvkiwskewfwvtdpskpfkcygfpvllaaahsnnyd 249
OY 245 CNRMFGDFTLSKSVNELLMEPOLKENSPEGASDVLLRYVPMKDIWIFRSCDLHLS 304
DB 250 ctnwlgdftllsksvndelwepkteqspgsidllqkyppecdiwifkscdfhn 309
OY 305 SVAIGNOEGKVYVMDLSCPPVILITKLSHNSQSVYRQTAMSVDSSTILACCEDGTIRMW 364
DB 310 qiaignregklywewgspvllarlynqgckspirqlavsfdsstllgagedgtlwrw 369
OY 365 DVI 367
DB 370 dev 372

RESULT 8

AAU00327
 ID AAU00327 standard; Protein; 378 AA.
 AC AAU00327;
 DT 12-SEP-2001 (first entry)
 DE Fertilisation-independent endosperm protein, ses2w.pk0015.b10.
 XX Fertilisation-independent endosperm; plant reproduction; apomixis;
 KM seed; pharmaceutical; nutraceutical; polymer; wheat; ses2w.pk0015.b10.
 OS Triticum aestivum.
 XX
 PN WO200116325-A2.
 PD 08-MAR-2001.
 PF 30-AUG-2000; 2000WO-US23735.
 PR 31-AUG-1999; 99US-0151575.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;
 PI Simmons CR, Weng Z;
 DR WPI: 2001-244407/25.
 DR N-PSDB; AAS01174.
 PT New plant fertilization independent endosperm protein for the
 PT production of seed without fertilization is recombinantly produced -
 XX
 PS Claim 1; Page 69-70; 94pp: English.
 XX
 CC The sequence represents the amino acid sequence of fertilisation-
 CC independent endosperm protein, encoded by cdna clone ses2w.pk0015.b10.
 CC Fertilisation-independent endosperm proteins are plant reproduction
 CC proteins necessary for apomixis, the formation of seeds without
 CC fertilisation. Apomixis is especially useful to agriculture because it
 CC eliminates the necessity of selfing plants to produce genetically
 CC identical seed. Such seed is useful to produce seeds for human and animal
 CC food and for commercial milling and extraction, including the production
 CC of useful recombinant products in the endosperm e.g. pharmaceutical,
 CC nutraceutical, industrial compounds and polymers. Embryoless seed
 CC production by transgenic plants is less likely to cause ethical and
 CC environmental concern over transgenic plant production as no gametes are
 CC being produced to cross pollinate with other crops and the seeds cannot
 CC germinate.
 XX
 SQ Sequence 378 AA;

Query Match 70.7%; Score 1394; DB 22; Length 378;
 Best Local Similarity 68.5%; Pred. No. 4.8e-146;
 Matches 248; Conservative 49; Mismatches 65; Indels 0; Gaps 0;

QY 6 LGNESIVSLPFSNKKSKVYNNRQEGKKPLVAVENFLDRFDEVYTAGNRTITLNC 65
 II : IIII : : : : : IIII : : : : : IIII : : : : : IIII : : : : :
 DB 10 lgcnaavslaprsrrekykthegkrpkyalgaifndarydvatvgngvtytyrg 69
 CC
 QY 66 LGDGAISALOSYADDEKSEFYTWAGVGNPNVYAGVKGITRVIDVNSSETIHKSLV 125
 I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
 DB 70 lpdgnlavlgayidaddqsfytlswacdldgtpllvaagnavlrvnatekikfksfl 129
 CC
 QY 126 GHGDSVNEIRTPQLKPOLVITASKDSYRLNWNETGICILIFAGAGGRHYELSDPHPS 185
 IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
 DB 130 ghgdsneirtpqlkpslifsaskdesvrlwnvhtgicillifagaggrhneylsvdfhps 189
 CC
 QY 186 DIYRASGCMPTTITINSMKEFTVVEKSFMTWDDSPKPFKTFVQFPPTASIHNNYDC 245
 IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
 DB 190 diyrtasgcmpttltvksmkeftvveksfwtwddlpskfpkftvgfplmtsvshnnydc 249

QY 246 NRMFGDILSKSDNEILLMEPOLKENSFGASDVLRLRPVPMCDIMFIRFSCDLHSS 305
 II : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
 DB 250 trwlgdfllsksvdneivlwepkkegpggsidvldkypdpdcidvfkfscdfhfrq 309
 QY 306 VAIGNOEGKRVYWDLKSCEPVLITRLSHNSKSVIRQTMASVDSSTIIACCDDGTIWRMD 365
 : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :
 DB 310 laignregklywewgasppvlltrlspsqcmprqtavsfldgstllacgedgslytwd 369
 QY 366 VI 367
 :
 DB 370 ev 371

RESULT 9
 AAU00325
 ID AAU00325 standard; Protein; 375 AA.
 XX
 AC AAU00325;
 DT 12-SEP-2001 (first entry)
 DE Fertilisation-independent endosperm protein, p0120.cdbdd48r.
 XX
 KM Fertilisation-independent endosperm; plant reproduction; apomixis;
 KM seed; pharmaceutical; nutraceutical; polymer; rice; p0120.cdbdd48r.
 XX
 OS Oryza sativa.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 279 /label= unknown
 FT /note= "Encoded by gan"
 XX
 PN WO200116325-A2.
 PD 08-MAR-2001.
 PF 30-AUG-2000; 2000WO-US23735.
 PR 31-AUG-1999; 99US-0151575.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;
 PI Simmons CR, Weng Z;
 DR WPI: 2001-244407/25.
 DR N-PSDB; AAS01172.
 PT New plant fertilization independent endosperm protein for the
 PT production of seed without fertilization is recombinantly produced -
 XX
 PS Claim 1; Page 65-67; 94pp: English.
 XX
 CC The sequence represents the amino acid sequence of fertilisation-
 CC independent endosperm protein, encoded by cdna clone p0120.cdbdd48r.
 CC Fertilisation-independent endosperm proteins are plant reproduction
 CC proteins necessary for apomixis, the formation of seeds without
 CC fertilisation. Apomixis is especially useful to agriculture because it
 CC eliminates the necessity of selfing plants to produce genetically
 CC identical seed. Such seed is useful to produce seeds for human and animal
 CC food and for commercial milling and extraction, including the production
 CC of useful recombinant products in the endosperm e.g. pharmaceutical,
 CC nutraceutical, industrial compounds and polymers. Embryoless seed
 CC production by transgenic plants is less likely to cause ethical and
 CC environmental concern over transgenic plant production as no gametes are
 CC being produced to cross pollinate with other crops and the seeds cannot
 CC germinate.
 XX
 SQ Sequence 375 AA;

Query Match 70.5%; Score 1391; DB 22; Length 375;
Best Local Similarity 69.2%; Pred. No. 1e-145;
Matches 245; Conservative 48; Mismatches 61; Indels 0; Gaps 0;

CC Fertilisation-independent endosperm proteins are plant reproduction
CC proteins necessary for apomixis, the formation of seeds without
CC fertilisation. Apomixis is especially useful to agriculture because it
CC eliminates the necessity of selfing plants to produce genetically
CC identical seed. Such seed is useful to produce seeds for human and animal
CC food and for commercial milling and extraction, including the production
CC of useful recombinant products in the endosperm e.g. pharmaceutical,
CC nutraceutical, industrial compounds and polymers. Embryoless seed
CC production by transgenic plants is less likely to cause ethical and
CC environmental concern over transgenic plant production as no gametes are
CC being produced to cross pollinate with other crops and the seeds cannot
CC germinate.

Sequence 391 AA:

Query Match 67.8%; Score 1337.5; DB 22; Length 391;
Best Local Similarity 70.7%; Pred. No. 9.9e-140;
Matches 239; Conservative 45; Mismatches 53; Indels 1; Gaps 1;

CC Fertilisation-independent endosperm protein, p0003.cgpded29rb.
KW seed; pharmaceutical; nutraceutical; polymer; p0003.cgpded29rb.
OS Zea mays.

RESULT 10
AAU00317
ID AAU00317 standard; Protein; 391 AA.
AC AAU00317;
XX
XX 12-SEP-2001 (first entry)
DE Fertilisation-independent endosperm protein, p0003.cgpded29rb.
XX
KW Fertilisation-independent endosperm; plant reproduction; apomixis;
KW seed; pharmaceutical; nutraceutical; polymer; p0003.cgpded29rb.
XX
OS Zea mays.

Key Location/Qualifiers
FT Misc-difference 340 /note= "Encoded by gc"
XX
XX WO200116325-A2.
XX
XX PD 08-MAR-2001.
XX
XX PF 30-AUG-2000; 2000WO-US23735.
XX
XX PR 31-AUG-1999; 99US-0151575.
XX
XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX PI Butler KH, DanilevsKaya O, Miao G, Morgante M, Sakai H;
XX Simmons CR, Weng Z;
XX
XX DR WPI: 2001-244407/25.
XX DR N-PSDB: AAS01164.
XX
XX PT New plant fertilization independent endosperm protein for the
XX production of seed without fertilization is recombinantly produced -
XX
XX PS Claim 1, page 52-53; 94pp; English.
XX
XX CC The sequence represents the amino acid sequence of fertilisation-
XX independent endosperm protein, encoded by cDNA clone p0003.cgpded29rb.
CC

CC Fertilisation-independent endosperm proteins are plant reproduction
CC proteins necessary for apomixis, the formation of seeds without
CC fertilisation. Apomixis is especially useful to agriculture because it
CC eliminates the necessity of selfing plants to produce genetically
CC identical seed. Such seed is useful to produce seeds for human and animal
CC food and for commercial milling and extraction, including the production
CC of useful recombinant products in the endosperm e.g. pharmaceutical,
CC nutraceutical, industrial compounds and polymers. Embryoless seed
CC production by transgenic plants is less likely to cause ethical and
CC environmental concern over transgenic plant production as no gametes are
CC being produced to cross pollinate with other crops and the seeds cannot
CC germinate.

Sequence 391 AA:

Query Match 67.8%; Score 1337.5; DB 22; Length 391;
Best Local Similarity 70.7%; Pred. No. 9.9e-140;
Matches 239; Conservative 45; Mismatches 53; Indels 1; Gaps 1;

CC Fertilisation-independent endosperm protein, p0003.cgpded29rb.
KW seed; pharmaceutical; nutraceutical; polymer; cen3npk0076.b8; corn.
OS Zea mays.

RESULT 11
AAU00311
ID AAU00311 standard; Protein; 461 AA.
AC AAU00311;
XX
XX 12-SEP-2001 (first entry)
DE Fertilisation-independent endosperm protein, cen3npk0076.b8.
XX
KW Fertilisation-independent endosperm; plant reproduction; apomixis;
KW seed; pharmaceutical; nutraceutical; polymer; cen3npk0076.b8; corn.
XX
OS Zea mays.

Key Location/Qualifiers
FT Misc-difference 340 /note= "Encoded by gc"
XX
XX WO200116325-A2.
XX
XX PD 08-MAR-2001.
XX
XX PF 30-AUG-2000; 2000WO-US23735.
XX
XX PR 31-AUG-1999; 99US-0151575.
XX
XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX PI Butler KH, DanilevsKaya O, Miao G, Morgante M, Sakai H;
XX Simmons CR, Weng Z;
XX
XX CC The sequence represents the amino acid sequence of fertilisation-
XX independent endosperm protein, encoded by cDNA clone p0003.cgpded29rb.
CC

DR	XX	WPI: 2001-244407/25.
DR	XX	N-PSDB: AAS01158.
PT	XX	New plant fertilization independent endosperm protein for the
PT	XX	production of seed without fertilization is recombinantly produced -
PS	XX	Claim 1; Page 42-44; 94pp: English.
CC	XX	The sequence represents the amino acid sequence of fertilisation-
CC	XX	independent endosperm protein, encoded by cDNA clone cen3npk0076.b8.
CC	XX	Fertilisation-independent endosperm proteins are plant reproduction
CC	XX	proteins necessary for apomixis, the formation of seeds without
CC	XX	fertilisation. Apomixis is especially useful to agriculture because it
CC	XX	eliminates the necessity of selfing plants to produce genetically
CC	XX	identical seed. Such seed is useful to produce seeds for human and animal
CC	XX	food and for commercial milling and extraction, including the production
CC	XX	of useful recombinant products in the endosperm e.g. pharmaceutical,
CC	XX	nutraceutical, industrial compounds and polymers. Embryotless seed
CC	XX	production by transgenic plants is less likely to cause ethical and
CC	XX	environmental concern over transgenic plant production as no gametes are
CC	XX	being produced to cross pollinate with other crops and the seeds cannot
CC	XX	germinate.
SO	XX	Sequence 461 AA;
Query Match		61.8%; Score 1219.5; DB 22; Length 461;
Best Local Similarity		59.0%; Pred. No. 1.7e-126;
Matches	214; Conservative	63; Mismatches 85; Indels 1; Gaps
OY	6	LGNESIVGSLTPSNKSKSYVTNRIQEKKRLVAVENFLDARFEDVFVTAAGNRITLNC 65
Db	68	Igcveevellvpsrkkrckreykpnskkytgnhbiyiaifnidmrydydvafalaacsvilyrc 127
OY	66	IGDGAISALOSTYAPDEKDESPFYTSVMACG-VNGNPVYAAGVGGIIRVIDVNSETIHKS L 124
Db	128	Ienggfllqnyvedkdksesfyllswrlidydspllvaaagsnrrilrvincateklaksl 187
OY	125	VGHGDSVNEIKTPOPLQPOLVTTASKDESVRILMNVETGICILIPAGAGHRYEVLSDVFP 184
Db	188	vghgsiheirthaskspelliisaekdesirlmwhtgicilvfagaagnrhvdlsvdfhp 247
OY	185	SDIYFAACGMDDTIKTISMKEFFTYVEKSFMTWDDESKRPFQEPVFTASIHNNYVD 244
Db	248	IevglfaacgmduetvkIwsmkefilyeksyswgphskprltnidfpvltaavhsdyvd 307
OY	245	CNRWFGRDLTLKSVNDNLIEWPDLKFNSPGASDVLLTFYPVPMCDIMEIFKRSCDLHS 304
Db	308	crrwfgrdlflksvynvaenallwekpdkrrrpgsgsvdvlqypvpkcselwmkscsflynn 367
OY	305	SVATINOGEGKYVWMDLKSCPVLITKLISHNOKSVINHQTAMSYDGSIIILACCDDGITRW 364
Db	368	qnaigmkqgelyvwegsppvlldirlcndgeckspirlgtavsdgftlllgaaaddgalwtw 427
OY	365	DVI 367
Db	428	dev 430
RESULT	12	
AAU00322	ID	AAU00322 standard; Protein: 461 AA.
XX	AAU00322;	
XX	12-SEP-2001	(first entry)
DE	Fertilisation-independent endosperm protein, p0104.cabbn62r.	
KW	Fertilisation-independent endosperm; plant reproduction; apomixis;	
KW	seed; pharmaceutical; nutraceutical; polymer; p0104.cabbn62r.	

OS	Zea mays
XX	wO200116325-72.
PN	08-MAR-2001.
XX	
PD	
XX	
PF	30-AUG-2000; 2000WO-US23735.
XX	
PR	31-AUG-1999; 99US-0151575.
XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E I.
XX	(PION-) PIONEER HI-BRED INT INC.
PI	Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;
PI	Stimmons CR, Meng Z;
DR	WPI; 2001-244407/25.
XX	N-PsDB; AAS01169.
PT	New plant fertilization independent endosperm protein for the
XX	production of seed without fertilization is recombinantly produced -
PS	Claim 1; Page 60-62; 94pp; English.
CC	The sequence represents the amino acid sequence of fertilisation-
CC	independent endosperm protein, encoded by cDNA clone p0104_cabnb62r.
CC	Fertilisation-independent endosperm proteins are plant reproduction
CC	proteins necessary for apomixis, the formation of seeds without
CC	fertilisation. Apomixis is especially useful to agriculture because it
CC	eliminates the necessity of selfing plants to produce genetically
CC	identical seed. Such seed is useful to produce seeds for human and animal
CC	food and for commercial milling and extraction, including the production
CC	of useful recombinant products in the endosperm e.g. pharmaceutical,
CC	nutraceutical, industrial compounds and polymers. Embryotest seed
CC	production by transgenic plants is less likely to cause ethical and
CC	environmental concern over transgenic plant production as no gametes are
CC	being produced to cross pollinate with other crops and the seeds cannot
XX	germinate.
XX	
SQ	Sequence 461 AA:
Query Match	61.8%; Score 1219.5; DB 22; Length 461;
Best Local Similarity	59.0%; Pred No.1.7e-126;
Matches 214; Conservative	63; Mismatches 85; Indels 1; Gaps 1;
OY	6 LGNESIVGSLTPESNKKSYRYVNRIOEGKRLXAVVENFLDAREFDVFYAGGNRTLYNC 65
Db	68 lgcvevegllvpsrkreykypnskyvtvgmnpriyafgfnidmydvfatascnsvlyrc 127
OY	66 LGGDAISALQSIADEDKESFYTWSACG-VNGNPVVAAGCYKGRIIRVDVNSETHIKL 124
Db	128 lengfgjllqnydckdesfylvlwtldgvdspllvaaagrnrlrvncateklcks1 187
OY	125 VHGDSVNEFRQPLKPOLVITASKCESYRLMNVENGICILIFPAGSGHREYLVSDFRP 184
Db	188 vghgsihetrhaaspsallssaskdesirlmvhgicilyfaagggnrhdvlsvdfrp 247
OY	185 SDIYRASAQMDDTIKIKSNMKFEWTVERSFTWTDDPSKEPTKEVOFPVTASIHNTYVD 244
Db	248 tcygfatasgmactviyamketfwlyveksyswtgpnskfptrndqfpvltlaavhadyd 307
OY	245 CNRMFSDFILSKSVDNELILMEPOLKENSPGCASVLYLRPPMKDMIFIRFSCDLIHS 304
Db	308 ctwlgldfilfskvaavllwebpdkrrpgsgsvdlqkyvpkcsllwfmkiscdfsn 367
OY	305 SVALGODEGVVYWDLKSCPPVLITKLSHNQSSVTRQTAMSVDSGTLLACCEDTGIWRM 364
Db	368 qmaignkngelyvweyqsspvlldrlcngeckspitrtavsdsgstcllgaadgaawtw 427
OY	365 DVI 367
Db	428 dev 430

RESULT 13
AAU00339 standard; Protein; 363 AA.
ID AAU00339 standard; Protein; 363 AA.
AC AAU00339;
XX
XX
XX 12-SEP-2001 (first entry)
DE Fertilisation-independent endosperm protein, p0107.cbca179r.
XX
XX Fertilisation-independent endosperm; plant reproduction; apomixis;
XX seed; pharmaceutical; nutraceutical; polymer; p0107.cbca179r. corn.
XX
OS Zea mays.
XX
XX WO200116325-A2.
XX
XX 08-MAR-2001.
XX
XX 30-AUG-2000; 2000WO-US23735.
XX
XX 31-AUG-1999; 99US-0151575.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
XX Simmons CR, Weng Z;
XX
XX WPI; 2001-244407/25.
XX
XX New plant fertilization independent endosperm protein for the
XX production of seed without fertilization is recombinantly produced -
XX
XX
XX Claim 1; Page 84-85; 94pp; English.
XX
XX The sequence represents the amino acid sequence of fertilisation-
XX independent endosperm protein, encoded by cDNA clone p0107.cbca179r.
XX Fertilisation-independent endosperm proteins are plant reproduction
XX proteins necessary for apomixis, the formation of seeds without
XX fertilisation. Apomixis is especially useful to agriculture because it
XX eliminates the necessity of selfing plants to produce genetically
XX identical seed. Such seed is useful to produce seeds for human and animal
XX food and for commercial milling and extraction, including the production
XX of useful recombinant products in the endosperm e.g. pharmaceutical,
XX nutraceutical, industrial compounds and polymers. Embryoless seed
XX production by transgenic plants is less likely to cause ethical and
XX environmental concern over transgenic plant production as no gametes are
XX being produced to cross pollinate with other crops and the seeds cannot
XX germinate.
XX Note: It is stated in the specification that this sequence is encoded by
XX the DNA sequence shown in AAS01186, however this is not the case.
XX
XX Sequence 363 AA:
SQ

Query Match 61.6%; Score 1215.5; DB 22; Length 363;
Best Local Similarity 58.7%; Pred. NO. 3.3e-126;
Matches 213; Conservative 63; Mismatches 86; Indels 1; Gaps 1;

QY 6 LGNESVSGSLTPSNKSKSYVTRIDEGKRPVLA VVFNFDARFDVYTAGCNRITLLNC 65
DB 1 lgcveevgllypsrkreykpskyltygnhplaylgnfidmryydvfatascsvilyrc 60
QY 66 LGDGAISALQSYADEDEKESFTVSWAGC-VNGNPPVAAAGVKGKIRIVDIVNSERTHKSL 124
DB 61 lenggflllgnvydekdesfytiswtidqvasspllvaaagnrlirvncatekidxel 120
QY 125 VGHGDSVNEIRTPQLKPOLVITASKDESVRLMNVETGICILIFAGAGGRHYEVLVDVPH 184
DB 121 vghggsiheltrhaskpsllisaskdesirlnvhticllvfaagqnhdvlsvdfnp 180

QY 185 SDIYRFASCGMDTTIKWGMKEFWYVEKSPFTWDDPSKFPTRKFOVPFVTAISHNNYD 244
DB 181 tevglifascgmdntvkiwmkewlyveksywtgpnskfiptrnqfpltaavhsdydv 240
QY 245 CNRMFGDFTLKSVNDIEILLMEPOLKENSPPGASDVLRLVPMKDIMFIRFSCDLHS 304
DB 241 ctrwlgdfllkskvknaevllwepkpkrrpgsgsvdlqkypkcsllwfmfscdfyan 300
QY 305 SVAIGNQEGKVVYMDLKSCEPPVLITKLSHNSKSVIRQITAMSYDGTIIACCEBDGTIMRW 364
DB 301 gmaignmkgeilywvsgspvlidrlcnqceksprlqtsavsfsgstlfgaaddgailwrw 360
QY 365 DVI 367
DB 361 dev 363

RESULT 14
AAU00318 standard; Protein; 273 AA.
ID AAU00318 standard; Protein; 273 AA.
AC AAU00318;
XX
XX
XX 12-SEP-2001 (first entry)
DE Fertilisation-independent endosperm protein, p0003.cgfn34rb.
XX
XX Fertilisation-independent endosperm; plant reproduction; apomixis;
XX seed; pharmaceutical; nutraceutical; polymer; p0003.cgfn34rb.
XX
XX
XX Zea mays.
XX
XX WO200116325-A2.
XX
XX 08-MAR-2001.
XX
XX 30-AUG-2000; 2000WO-US23735.
XX
XX 31-AUG-1999; 99US-0151575.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
XX Simmons CR, Weng Z;
XX
XX WPI; 2001-244407/25.
XX N-PSDB; AAS01165.
XX
XX New plant fertilization independent endosperm protein for the
XX production of seed without fertilization is recombinantly produced -
XX
XX
XX Claim 1; Page 54-55; 94pp; English.
XX
XX The sequence represents the amino acid sequence of fertilisation-
XX independent endosperm protein, encoded by cDNA clone p0003.cgfn34rb.
XX Fertilisation-independent endosperm proteins are plant reproduction
XX proteins necessary for apomixis, the formation of seeds without
XX fertilisation. Apomixis is especially useful to agriculture because it
XX eliminates the necessity of selfing plants to produce genetically
XX identical seed. Such seed is useful to produce seeds for human and animal
XX food and for commercial milling and extraction, including the production
XX of useful recombinant products in the endosperm e.g. pharmaceutical,
XX nutraceutical, industrial compounds and polymers. Embryoless seed
XX production by transgenic plants is less likely to cause ethical and
XX environmental concern over transgenic plant production as no gametes are
XX being produced to cross pollinate with other crops and the seeds cannot
XX germinate.
XX
XX Sequence 273 AA:
SQ

Query Match 55.98; Score 1103; DB 22; Length 273;
Best Local Similarity 74.38; Pred. No. 7e-114;
Matches 197; Conservative 30; Mismatches 38; Indels 0; Gaps 0;
QY 103 AGGKGIIRVDVNSFTIHKSLVGHGDSVNEIRTPOLKPOLVITASDESVRLMNVENG 162
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 2 asgnglirvncateklaksfyvghgdsineirtpkpslilsasdesvrlmnhgql 61
QY 163 CILIFAGAGHRYEVLVSDPHPSDIYRFASCGMDTTIKIMSKFEFTYVEKSFMTDPS 222
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 62 cillifagagrhnevlsdvdfpsdierfasgmdntvkiwsketwlyvdkyswtldps 121
QY 223 KFPKRYVQFPYFTASIRTNVYDCNRMGDFILSKSVNDNEILMEPOLKENSPEGASDYL 282
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 122 kfpkyvqfpylilaavsnvdcrtwlgdflilsksvndeiylwepkckegspgsgsldll 181
QY 283 LRYVPMCDIWFIFESDILHSSVAIGNOEGKYVWMDKSPVLIITKLSHNSKSVIRQ 342
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 182 qkyvpedlwfikscdfhnglaignregklywvqspvllarllynqgckspirg 241
QY 343 TAMSVDSTILACCEDGTWRMDVI 367
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 242 tawfdgstllgagedgtlwrwdev 266
RESULT 15
AAU00320
ID AAU00320 standard; Protein; 430 AA.
XX
AC AAU00320;
XX
DT 12-SEP-2001 (first entry)
DE Fertilisation-independent endosperm protein, p0041.crtaw93r.
XX
KM Fertilisation-independent endosperm; plant reproduction; apomixis;
KW seed; pharmaceutical; nutraceutical; polymer; p0041.crtaw93r.
XX
OS Zea mays.
XX
FH Key location/Qualifiers
FT Misc-difference 322 /note= "Encoded by tgtt"
XX
PN W0200116325-A2.
XX
PD 08-MAR-2001.
XX
PF 30-AUG-2000; 2000WO-US23735.
XX
PR 31-AUG-1999; 99US-0151575.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Butler KH, Danilevska O, Miao G, Morgante M, Sakai H;
XX
PI Simmons CR, Weng Z;
XX
DR WPI: 2001-244407/25.
XX
DR N-PSDB: AAS01167.
XX
PT New plant fertilization independent endosperm protein for the
XX
XX production of seed without fertilization is recombinantly produced -
PS Claim 1; Page 56-57; 94pp; English.
XX
XX The sequence represents the amino acid sequence of fertilisation-
XX independent endosperm protein, encoded by cDNA clone p0041.crtaw93r.
XX Fertilisation-independent endosperm proteins are plant reproduction
XX proteins necessary for apomixis, the formation of seeds without
XX fertilisation. Apomixis is especially useful to agriculture because it
XX eliminates the necessity of selling plants to produce genetically
XX identical seed. Such seed is useful to produce seeds for human and animal

CC food and for commercial milling and extraction, including the production
CC of useful recombinant products in the endosperm e.g. pharmaceutical,
CC nutraceutical, industrial compounds and polymers. Embryoless seed
CC production by transgenic plants is less likely to cause ethical and
CC environmental concern over transgenic plant production as no gametes are
CC being produced to cross pollinate with other crops and the seeds cannot
CC germinate.
XX
SQ Sequence 430 AA;
Query Match 53.08; Score 1045; DB 22; Length 430;
Best Local Similarity 51.88; Pred. No. 4e-107;
Matches 188; Conservative 59; Mismatches 84; Indels 32; Gaps 2;
QY 6 LGNESIVGSLTPPSNKKSKYVTRIOEGKKPLVAVVFNFLDARFPDVPVTFAGNRTLYNC 65
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 68 lgcvevvegllvpsrkreykpnskytvgnhplayalgnfidmryydvfalscnsvilyrc 127
QY 66 LGDGAISALOSYADEDEKEESFTYVSWAG-VGNPYYAAGGVKGIIRVDVNSFTIHKSL 124
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 128 lenggfgllqnyvdedkdesfytlswtidqvdspllvaagsnrlirvlnocateklkls 187
QY 125 VGHGDSVNEIRTPOLKPOLVITASKDSVRLMNVETICILIFAGAGHRYEVLVSDPHR 184
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 188 vghgdsineirtpolkpolvitasksdvrlmnaveticiilifagaghrvylsvdphr 216
QY 185 SDIYRFASCGMDTTIKIMSKFEFTYVEKSFMTDPSKEPTKFPVPPFTASIRTNVVD 244
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 217 tevgifascgmdntvkiwsketwlyvdkyswtgshpskfpitriqfpylilaavhsdyvd 276
QY 245 CNRMFGDFILSKSVNDNEILMEPOLKENSPEGASDYLKRYVPMCDIWFIFESCDILHS 304
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 277 ctrwlgdflilsksvndeiylwepkckirrpgegsvdvlgkypvpckhlwfmkfcdfysn 336
QY 305 SVALGNDBGKYVWDLKSCPPVLIITKLSHNSKSVIRIQTASVDSGFTILACCEDGTWR 364
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 337 qmalignkgeilywvqspvllidrlcngckspirgtaavsdgscilgaaddgalwv 396
QY 365 DVI 367
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 397 dev 399

Search completed: June 25, 2002, 01:13:53
Job time: 3737 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 00:17:26 ; Search time 73.23 Seconds

(without alignments)
484.186 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973

Sequence: 1 MSKITLGNESTVGLTPSNK.....STILACCEDGTIRWMDVITK 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	36.2	425	2	S58672
2	217.5	11.0	1189	2	AI2493
3	212.5	10.8	332	2	T29827
4	211.5	10.7	701	2	T16607
5	211.5	10.7	1227	2	AE1810
6	210.5	10.7	1356	2	T18521
7	202.5	10.3	704	2	S3263
8	202	10.2	559	2	AB2202
9	197.5	10.0	920	2	C96831
10	195	9.9	1526	2	AC2239
11	194.5	9.9	926	2	G96563
12	194.5	9.9	1747	2	AC1842
13	193	9.8	1258	2	AI2155
14	188	9.5	798	2	S34023
15	187.5	9.5	421	2	T14773
16	183.5	9.3	586	2	T38992
17	183	9.3	1711	2	AD1842
18	182	9.2	606	2	T08180
19	181.5	9.2	380	2	T02085
20	181	9.2	1683	2	AF2071
21	179.5	9.1	518	2	B48088
22	176	8.9	377	2	T07376
23	176	8.9	616	2	C86239
24	175.5	8.9	1146	2	A5532
25	174.5	8.8	876	2	T51507
26	172.5	8.7	380	2	T03765
27	172	8.7	934	2	AG1889
28	170.5	8.6	515	2	S19487
29	170	8.6	377	2	T04089

30	169.5	8.6	304	2	AG1837	WD-40 repeat prote
31	169	8.6	703	2	T43557	F-box/WD-repeat pr
32	168.5	8.5	376	2	T19266	hypothetical prote
33	168	8.5	377	2	T04086	GTP binding protei
34	168	8.5	409	2	S36113	lts-1 protein - hu
35	168	8.5	410	2	S48052	platelet-activatin
36	166.5	8.4	1551	2	AB2410	WD-repeat protein
37	166	8.4	375	2	T03256	GTP-binding protei
38	166	8.4	502	2	T41148	trp-asp repeat con
39	165	8.4	310	2	T43158	probable GTP-bind
40	164.5	8.3	326	2	T16987	GTP-binding protei
41	164.5	8.3	494	2	T19550	hypothetical prote
42	163.5	8.3	642	2	T39490	transcription init
43	163	8.3	713	2	JN0133	WD-40 repeat regul
44	162.5	8.2	240	2	T02059	GTP-binding regula
45	162.5	8.2	906	2	S35342	Golgi-associated p

ALIGNMENTS

RESULT 1
S58672
extra sex combs protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 15-Feb-1996 #sequence-revision 01-Mar-1996 #text-change 26-May-2000
R:Gutjahr, T.; Freil, E.; Spicer, C.; Baumgartner, S.; White, R.A.H.; Noll, M.
EMBO J. 14, 4296-4306, 1995
A:Title: The polycomb-group gene, extra sex combs, encodes a nuclear member of the WD
A:Reference number: S58672; MUID:96016202
A:Accession: S58672
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-425 <GUT>
A:Cross-references: GB:141867; NID:g1050996; PIDN:AAA86427.1; PID:g1050997
C:Genetics:
A:Gene: FlyBase:esc
A:Cross-references: FlyBase:FBgn0000588
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match	Best Local Similarity	36.2%	Score 715;	DB 2;	Length 425;
Matches 143;	Conservative 76;	Mismatches 129;	Indels 16;	Gaps 10;	
QY	17	PSNRKSKYKYNRIQEGK-KPLVAVNFELDARFF-DVFTAGGNRTILYNCIGDAISAL	74		
DB	56	PKSRAYKXKDTYHKENHGANIFGVAFNTLLGKDEPQVFATAGSNRYTYECPRQGMQLL	115		
QY	75	QSTADEKESFYTVSMAGV-NGNPYVAAAGYKGIIRIVIDVNSEFTIHSLSVGHGDSVNE	133		
DB	116	HCYADDPDEVEFTYCAMSYDLKTSPLAAAGRGVIRIVDVQNRVAVNYIGHGQAIN	175		
QY	134	IRNPQLPOLVITASKDESRMLNVTGICILFACAGRGYVLSVDFPHSPDIYFASG	193		
DB	176	LKPHPKLQLLSGSKDHAIRLNNIQSHVCITALLGVGEGRDVLSIDFMRG-DRIVSS	234		
QY	194	GMPYTKIKISMK--EFWTVYKSFYTWDDPS--KPTKFEQPFV--TASHTNYVVCNRM	248		
DB	235	GMHSLAKLMCLNTPFERHKKIELSNFISQEKSTLPFTYVKKHFDFTROIRHRYVVCVOM	294		
QY	249	FQDFILSKSVNDEILMER-QIKEN---SPGASDVLLRYVPKCDIWFIFKSCDLH	303		
DB	295	FGNFVLSKSCENALVCMKPGQLHQSFEQVPPSDSCITIAEFYDCEIWFVAFGPNPQ	354		
QY	304	SSVAITNOEGKYVMDLKSPP--VLITKLSHNSQSVIRQTFMSVDSGTIACCGDGTI	361		
DB	355	KVALTNOEGKYVMDLPSDPEGAMHTLLHNSRVAIVKQIAFSKDSAVLYVCDATV	414		
QY	362	WRWD 365			
DB	415	WRWN 418			

Db 409 -----DTLEF-----VRLAGHRGACLOQRGLVWSSGSDNTIRLMDIH---- 449
 QY 273 SPGEASDVLLRY-----PVPMDIWFIFKSCDLHLSVAIGNQEKVYVMDLKS--PP 325
 Db 450 -----SGCYLRYLEGHHELVRC-----IRFD-----EKRIYSGAYDGKIKVMDLQALDPR 495
 QY 326 VLTIKLSHNSKSVIRQTA-----MSVDSSTILACCEGDTIWRMDVI 367
 Db 496 ALSSEIC-----LCSLVQHTGRVFRLOQFDPOIYSSSHDITLIMDFL 538

RESULT 5

AE1810
 WP-40 repeat protein [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AE1810
 R:Kaneke, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriguchi, Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AE1810
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1227 <KRP>
 A:Cross-references: GB:BA000019; PIDN:BA077553.1; PID:g17135007; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr0029

Query Match 10.7%; Score 211.5; DB 2; Length 1227;
 Best Local Similarity 25.1%; Pred. No. 2.3e-09;
 Matches 82; Conservative 50; Mismatches 114; Indels 81; Gaps 15;

QY 68 DGAISALQSVADDEKESFYVTS-----WACGVNCP-----YVAGGVK 107
 Db 697 DGRILA---SASDQTIKIMDITGNCQOTLIGHDWMVSVTFPTVTDPRILLASSAD 753
 QY 108 GIIRIVDNSETIHKSLVGHGDSVNEIRTOPLKPOLVITASKDESVRLNVENTGICILAF 167
 Db 754 QHKLWDVATGCKLTAKGHTREVHSVFSF--DQOTLASSGDSYRLMDVKQGCQWIF 812
 QY 168 AGAGGRVYLVSDPHPSDIYRASCAGMOTITIKMSKE-----FWTYEAKSFYTTDDP 221
 Db 813 E---GHSKRVYSVRFSP--DQOTLASSGDSRSTIKLMDIORCECVNTLMGHSSQYWAIAFSP 868
 QY 222 SKFPFTFVGFVPTATSIHTNYVDCNRMFGLTILSKSVNDELLIMEPOLKENSPEGASDV 281
 Db 869 D-----GRTILSCSDDOTARLMD-VITGNS-----LNI 895
 QY 282 LRRYPVPMCDIMFIFKSCDLHLSVAIGNQEKVYVMDLKS--CPPVLITIKLSHNSKSV 339
 Db 896 LRGY---TRDYVSVAFSPPSQI---LASGRDIDYIIGMLNKTGCCHP-----LNGHGR-- 943
 QY 340 IRTQAMSVDSSTILACCEGDTIWRMDV 366
 Db 944 IRSVAFHPDGKILASSGSDNTIRLMDI 970

RESULT 6

AE18521
 beta transducin-like protein - Podospora anserina

C:Species: Podospora anserina
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18521
 R:Saune, S.; Turcq, B.; Begueret, J.
 Gene 162, 135-139, 1995
 A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina
 A:Reference number: Z18944; MUID:96009891
 A:Accession: T18521

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1356 <SAU>
 A:Cross-references: EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775.1
 C:Genetics:
 A:gene: het-el
 A:introns: 761/3

Query Match 10.7%; Score 210.5; DB 2; Length 1356;
 Best Local Similarity 28.2%; Pred. No. 3.2e-09;
 Matches 78; Conservative 31; Mismatches 103; Indels 65; Gaps 11;

QY 101 VAAGVKGIIIRIVDNSETIHKSLVGHGDSVNEIRTOPLKPOLVITASKDESVRLNVENT 160
 Db 940 VASGSDDHITIKMDASGTOITLEGHSSVLSVAFSP--DQGVASGSDGDKTIKMDTPAS 998
 QY 161 GICILIFAGAGHRYEVLSDPHPSDIYRASCAGMOTITIKMSKEFWYVEKSFYTTDD 220
 Db 999 GTCOTGLEHSG--SVSVAFSP--DQGVASGSDDKTIKIM-----DT 1038
 QY 221 PSKFPFK-----FYQFPYFTASIHNTNYVDCNRMFGDFILSKSVNDELLMEP---OLK 270
 Db 1039 ASGCTOTGLEHSGVQSVVFSFSP-----GQVVASGSDHITIKIMDAVSGCT 1086
 QY 271 ENSPGEASDVLLRYPPMCDIWFIFKSCDLHLSVAIGNQEKVYVMDLKS--CPPVLI 328
 Db 1087 QLEGHGDS-----VMSVAFSP--GQVVASGSDITIKIMDAAGTCTOTL- 1131
 QY 329 TKLSHNSKSVIRQTAHSVDSSTILACCEGDTIWRMD 365
 Db 1132 -----EGHGGVHSAVAFSPDQGVASGSDITIKIMD 1163

RESULT 7

transcription initiation factor IID-associated protein, 80K - fruit fly (Drosophila m
 N:Alternate names: tightly-associated factor TAF-II-80

C:Species: Drosophila melanogaster
 C:Date: 31-Mar-1992 #sequence_revision 14-Sep-1994 #text_change 01-Dec-2000
 C:Accession: S33263; S45729; A54593
 R:Dynlacht, B.D.; Weinzierl, R.O.J.; Admon, A.; Tjian, R.
 Nature 363, 176-179, 1993
 A:Title: The DRAF(II)80 subunit of Drosophila TFIID contains beta-transducin repeats.
 A:Reference number: S33263; MUID:93247643
 A:Accession: S33263
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-704 <DYN>
 A:Accession: S45729
 A:Molecule type: protein
 A:Residues: 105-126;226-268;273-301;321-336 <DYZ>
 R:Fokkudo, T.; Gong, D.W.; Yamashita, S.; Takada, R.; Roeder, R.G.; Horikoshi, M.; Nak
 Mol. Cell. Biol. 13, 7859-7863, 1993
 A:Title: Molecular cloning, expression, and characterization of the Drosophila 85-kil
 A:Reference number: A54593; MUID:94067146
 A:Contents: Schneider cells, embryos
 A:Accession: A54593
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid; protein
 A:Residues: 1-409; 'KL' 412-704 <KOK>
 A:Note: sequence extracted from NCBI backbone (NCBIP.139815)
 C:Genetics:
 A:gene: FlyBase:Taf80
 A:Cross-references: FlyBase:FBgn0010356
 A:Map position: 47C5-6
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 C:Keywords: transcription initiation
 F:374-406/Domain: WD repeat homology <WD1>
 F:446-479/Domain: WD repeat homology <WD2>
 F:488-521/Domain: WD repeat homology <WD3>
 F:530-563/Domain: WD repeat homology <WD4>
 F:572-605/Domain: WD repeat homology <WD5>

	Query Match	9.9%	Score 195;	DB 2:	Length 1526;
	Best Local Similarity	26.6%;	Pred. No. 7.6e-08;		
	Matches 79;	Conservative 37;	Mismatches 107;	Indels 74;	Gaps 14;
Oy	90	SMACVGNCPN--YAAAGVGKGIIRVIDNSEFTIHKSLVGHGDSVNELRITQELPKRQLVIT	146		
Dd	1243	SWNSVVENPDGSMLASGSSDKTYVLWMDISSCKCLHTFOGHNNWNVAFN-DGSMLAS	1301		
Oy	147	ASKDESVRLMNVETGICILIFAGAGHRYEVLSDPHESDIYRFASCCMDTIIKIWSMKE	206		
Dd	1302	GSQDQTVRLMEISSLCKLHTFQ---GHTSWVSVPFSF-DGMLASGSDQDTVRLMSTSS	1357		
Oy	207	--FWTYVEKSFYTWDPSKFPFKVFQPFVTASLIHTNYVDONRMFGDFILSK-----	256		
Dd	1358	GECLTYTFIG-----HTN-----WVGSIFFSPDGAILLAS	1385		
Oy	257	-SYDNEILLMPOLKENSGEGASDVLLRPYPKCDIWF--IKESCDJLHLSVAIGNOG	313		
Dd	1386	GSQDQTVRLM-----SISGSKCLYTLQGHNHWGVSIFFSPDGTL--LASGSDQ	1432		
Oy	314	KYYVNDLKS--CPVLTIRKLSHNOSSVIROTAMSGVDSSTILAACCEDGITMMWDVIT	368		
Dd	1433	TYRLMNISSGEC---LYTLHGHIINS---VRSVAEFSDDGLILASGSDDETETIKLMDVKT	1483		

RESULT 11
 G96563
 Probable coatomer complex subunit, 33791-27676 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: G96563
 R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
 Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, T.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
 C.A.; Liu, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzialis
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141, MUID:21016719
 A:Accession: G96563
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1..926 <STD>
 A:Cross-references: GB:AE005173; NID:g10645445; PIDN:AMG21561.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F19K6.16
 A:Map position: 1
 A:Superfamily: coatomer complex beta' chain; WD repeat homology

	Query Match	9.9%	Score 194.5:	DB 2:	Length 926:
	Best Local Similarity	26.2%:	Pred. No. 4.2e-08:		
	Matches	60;	Conservative	35;	Mismatches 81; Indels 53; Gaps 9.
OY	100 YVAGGVKGIIRVIDYNSETIHKSLVGHDSVNEIRTPOLKPOLYITASKDESVRLMNYE	159	:	:	:
	: :	:	:	:	:
Db	71 WVAAGADMYIVRYNKTMDKYKFEAAHS DYTRCVAVHPLE-YVLSSDDMLIKLME	129	:	:	:
	: :	:	:	:	:
OY	160 TG-TGILLFAGAGGHREYLELVDFHPSDLYRRASCGMTPTIKINMSKEPWTVEKSFTYT	218	:	:	:
	: :	:	:	:	:
Db	130 KMACTQLE--GSHSYVMQVTENPKDINTPFASASLDPTIKILWN-----GS	174	:	:	:
	: :	:	:	:	:
OY	219 DDPSEPTPEKVOPEVFETASIHNTNYDCNMFW--GG--FLTSKVDNELIMEPOLKE---	271	:	:	:
	: :	:	:	:	:
Db	175 PDPN-----FTLDAHQGKNVCVDYFLFGGDKPYLTISDSDHTAKVMPDYQRKSCVQ	223	:	:	:
	: :	:	:	:	:
OY	272 --NSGEGASDYLKRPVPMCDIWITFKSCLHLSSVALIGNOGEKKYYW	318	:	:	:
	: :	:	:	:	:
Db	224 TLEGTHNNVASCFHEPLPI-----IIIGSDGDGYRIW	256	:	:	:
	: :	:	:	:	:

RESULT 12

WD-40 repeat protein [imported] - Anabaena sp. (strain PCC 7120)
AC1842

C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC1842

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriig
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat
DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC1842

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <R>
A:Cross-references: GB:BA000019; PIDN:BAB77807.1; PID:gi17135261; GSPDB:GNO0179

A:Experimental source: strain PCC 7120

C:Genetics:
A:Gene: al10283

[illegible]

RESULT 13

AT2155

WD-repeat protein [imported] - *Anabaena* sp. (strain PCC 7120)

C:Species: *Anabaena* sp.

A:Note: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120

C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C:accession: AT2155

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Irigoin, J.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 8, 205-213, 2001

A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AT2155

A:status: preliminary

A:molecule type: DNA

A:Residues: 1-1258 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA074499.1; PID:g17j31893; GSPDB:GN00179

A: Molecule type: protein

Query Match	9.5
-------------	-----

Query Match	9.5
-------------	-----

Query Match	9.5
-------------	-----

Best Local Similarity 21.2%; Pred. No. 5.7e-08;
Matches 83; Conservative 66; Mismatches 130; Indels 113; Gaps 19;

```
QY 6 LGNESTVGLTPSNKKSYVTNRIOEG---KKPLAVNFNLDARFEDVFTAGNRITL 62
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 67 LQNFLLHGTKNVTNSSLRLPRONSDGGQKNRP-----REHI 103
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 63 YNCLGGGALSAQSYADEKKESEFTYVSW-----ACVGNPNPYVAAGVKGTI 110
   | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 104 IDC-GDIWVSLAFGSSVPEKQSRVCYNIEMHREFREGODQLLATGLN-----NGRI 152
   | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 111 RYIDVNSEFTIHKSLYGHGDSVNEIRTPQLKPOLVITASKDESVRLMNV-ETGICILIFAG 169
   : | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 153 KIMDYVTGKLLNLVNDHTEVNRDLTFARPDSLLIVASRDKTLRVWDLKDDGNMKVLR- 211
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 170 AGGHRTEVLSDPHESDIYRFASCGMDTTIKIWSMKEFWTYEKSFTWTDDPSKPTKPV 229
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 212 --GHQWVYSCAFSP--DSSMLCSVGASKAVFLNMMDKY-TMIRK----- 251
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 230 QPVPVFTASIHITNVVDCNRWF---GDFILSKSVNELLIMEPOLKENSPEGASDVLLRY- 285
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 252 -----LEGHHHDVVACD--FSPDGALLATASYDFRYIMDPH-----NGDILMEFG 295
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 286 -----PVPKC-----DIWF--IKFSCD-LHLSVAIGNQEGKYVMDLKSPPVLITKLS 332
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 296 HLEPPPTPIFAGGANDRWVRYSVSFSDHGLHVASLA---DDKMYRFWRIDEDYPVQVAPLS 352
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 333 HNOSSKVINROTAMSVDSGTLLACCEDEGTIMRW 364
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 353 NGIC-----CAFSSTDGSVLAAGTHDGSYYFW 378
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: June 25, 2002, 01:16:42
Job time: 3556 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 01:15:16 ; Search time 37.87 Seconds
(without alignments)
377.278 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973
Sequence: 1 MSKITLGNESIVGSLTPSNK.....STLACCEDGTIMRWIVTK 369

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	211.5	10.7	665	1	L123_CAEEL
2	210.5	10.7	1356	1	HEP1_PODAN
3	202.5	10.3	704	1	T2D4_DROME
4	194.5	9.9	542	1	FW1B_HUMAN
5	193	9.8	1258	1	YS00_ANASP
6	188	9.5	798	1	T2D4_YEAST
7	188	9.5	914	1	COPP_DROME
8	183.5	9.3	586	1	TU12_SCHPO
9	182	9.2	606	1	PE20_CHLRE
10	181.5	9.2	380	1	GBB_MAIZE
11	179.5	9.1	518	1	TRCB_XENLA
12	177.5	9.0	605	1	FW1A_HUMAN
13	176	8.9	377	1	GBB_SOLTU
14	176	8.9	742	1	KWMA_THECU
15	175.5	8.9	1146	1	KWMA_DICDI
16	175	8.9	316	1	GBLP_BIOGL
17	173	8.8	1249	1	APAF_RAT
18	172.5	8.7	380	1	GBB_ORYSA
19	171	8.7	318	1	GBLP_DROME
20	170.5	8.6	515	1	YCW2_YEAST
21	170	8.6	377	1	GBB2_TOBAC
22	169.5	8.6	318	1	GBLP_TRYBR
23	169	8.6	703	1	POP2_SCHPO
24	168.5	8.5	376	1	YK4_CAEEL
25	168.5	8.5	904	1	COPP_RAT
26	168	8.5	377	1	GBB1_TOBAC
27	168	8.5	409	1	L1S1_BOVIN
28	168	8.5	409	1	L1S1_HUMAN
29	168	8.5	409	1	L1S1_MOUSE
30	167.5	8.5	800	1	T2D4_HUMAN
31	166	8.4	375	1	GBB3_TOBAC
32	166	8.4	1249	1	APAF_MOUSE
33	165.5	8.4	514	1	TUPL_CANAL

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description
1	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
2	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
3	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
4	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
5	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
6	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
7	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
8	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
9	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
10	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
11	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
12	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
13	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
14	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
15	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
16	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
17	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
18	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
19	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
20	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
21	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
22	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
23	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
24	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
25	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
26	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
27	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
28	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
29	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
30	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
31	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
32	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
33	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
34	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
35	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
36	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
37	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
38	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
39	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
40	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
41	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
42	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
43	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
44	L123_CAEEL	1973	10.7	665	1	L123_CAEEL
45	L123_CAEEL	1973	10.7	665	1	L123_CAEEL

DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Transcription initiation factor TFIID 85 kDa subunit (P85) (TAFII-80).
 GN TAF80 OR CG7704.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Ephyridoptera; Diptera; Brachycera; Muscomorpha;
 CC Ephyridoptera; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 105-120, 272-280 AND 365-374.
 RC TISSUE=Embryo;
 RX MEDLINE=94067146; PubMed=8247000;
 RA Koriya T., Gong D.-W., Yamashita S., Takada R., Roeder R.G.,
 RA Horikoshi M., Nakatani Y.;
 RT "Molecular cloning, expression, and characterization of the
 RT Drosophila 85-kilodalton TFIID subunit.";
 RL Mol. Cell. Biol. 13:7859-7863(1993).
 [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=93247643; PubMed=8483503;
 RA Dynlacht B.D., Weinzierl R.O.J., Admon A., Tjian R.;
 RT "The TAFII80 subunit of Drosophila TFIID contains beta-transducin
 RT repeats.";
 RL Nature 363:176-179(1993).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J., H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
 RA Beeson K.Y., Benson P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butler L.C., Busam D.A., Butler H., Cadenat E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Flier C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paule J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [4]
 RP FUNCTION: TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
 RP CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
 RP AND REPRESSORS. TAFII-80 INTERACTS DIRECTLY WITH TBP AND TAFII-110.
 RP MAY PLAY A ROLE IN HELPING TO ANCHOR TAFII-110 WITHIN THE TFIID
 RP COMPLEX. MAY BE INVOLVED IN TRANSDUCING SIGNALS FROM VARIOUS

CC TRANSCRIPTIONAL REGULATORS TO THE RNA POLYMERASE II TRANSCRIPTION
 CC MACHINERY.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 CC -1- SIMILARITY: TO YEAST TAFII-90 AND MAMMALIAN TAF2D.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U06460; AAC6481.1; -;
 CC EMBL: AE003828; AAF58737.1; -;
 CC TRANSFAC: T02122; -;
 CC DR FLYBASE: FBgn0010356; TAF80.
 CC DR InterPro: IPR001680; WD40.
 CC DR Pfam: PF00400; WD40; 6.
 CC DR PRINTS: PR00320; GPROTEINBRPT.
 CC DR SMART: SM00320; WD40; 6.
 CC DR PROSITE: PS00678; WD_REPEATS_1; 3.
 CC DR PROSITE: PS50082; WD_REPEATS_2; 6.
 CC DR PROSITE: PS50294; WD_REPEATS_REGION; 2.
 CC DR Transcription regulation: Nuclear protein; Repeat; WD repeat.
 CC KW REPEAT
 CC FT REPEAT 376 405 WD 1.
 CC FT REPEAT 448 478 WD 2.
 CC FT REPEAT 490 520 WD 3.
 CC FT REPEAT 532 562 WD 4.
 CC FT REPEAT 574 604 WD 5.
 CC FT REPEAT 616 646 WD 6.
 CC FT CONFLICT 410 411 KL -> NV (IN REF. 2).
 CC SQ SEQUENCE 704 AA; 79324 MW; F4E2F12534501EF6 CRC64;
 Query Match 10.3%; Score 202.5; DB 1; Length 704;
 Best Local Similarity 25.2%; Pred. No. 2.2e-09;
 Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;
 QY 110 IIVDIIVNSITKHSVGHSDSVNEIRTOPLKPOLVTTASKDSVRLMNVETGICILIFAG 169
 DB 432 VMLDNRSEVTRSLMGHTGPRYRCAP-EMNLLISCESDSTIKMLSLTSCVYTYR- 489
 QY 170 AGHRYEVLVDVPHSDIYRFASCGMDTITIKISMKEFVTYVEKSFVTWDDPSKEPTKEV 229
 DB 490 --GHYVPVWDVAFAPRGY- FVSCSYDKTARLMA-----TDSNQ----- 525
 QY 230 QRPVPLASHITHTVDCNRKF--GDPLTKSVNELLMPOLKENSFGCAGADVLIRPV 287
 DB 526 ALRVFVG--HLSDVDVCPHPNSNVATGSSDRTYRLW-----DNMTGQSVR-LMTGHNG 577
 QY 288 PKMDIIFIKESCDLHSAIGNOEKGYVMDLKSCPPVLITKLHNSQKSVYRTAMSV 347
 DB 578 SVSSLAFL--SAGGRYLA---GSVDHNIITMDL-SNGSLVTLIRHT---SVYITITFSR 628
 QY 348 DSGTILACCEDEGTIRWD 365
 DB 629 DGTVLAAGLDNNLTFLMD 646
 RESULT 4
 FW1B_HUMAN STANDARD; PRT; 542 AA.
 AC 09UKB1: 09Y4C6; 09P2S8; 09P2S9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE F-box/wd-repeat protein 1B (F-box and wd-repeats protein beta-Trcp2).
 GN FBXW1B OR FBW1B OR BTRCP2 OR KIAA0696.
 OS Homo sapiens (Human).

FT	VARSPIC	16	48	CSVSRISMGICANIVSMSCALCSOMPVSCL -> NTSV
FT	SEQUENCE	542 AA;	62090 MW;	MEDONEDBSPKRNITM (IN ISOFORM B).
SO				7CD40087EFAA5C84 CRC64;
Query Match		9.9%;	Score 194.5; DB 1;	Length 542;
Best Local Similarity		26.1%;	Pred. No. 7,3e-09;	
Matches 77;	Conservative	44;	Mismatches 107;	Indels 67; Gaps 14;
QY	87	YTVSMACGVNCPNPNVAAAGVGIIRIVIDVNSETTHKSLVGHGDSVNEIRTPDLKQOLVIT	146	
Db	279	HTGSVLCLQGYDERIAYVSSSDSTVAKWMDVNGEVLNLTILHNEAVALHRLR---FSNGLMVT	335	
QY	147	ASKEESVRLMNVENGICILIFAGAGHRYEVLVSDFHSDIYRFASCGMDTIKIKMSKE	206	
Db	336	CSKRSLIAVWDMASATDTTLRLRVLVGHRAAVVWDF--DDKIVASG-DRTIKVMS---	389	
QY	207	FMTYVESFTWTDPSKPTKFPDPPEVTASIHNTVYDCNRMFGDFILSKSYDNELLME	266	
Db	390	-----TSTCE-VTILNKHKGIACTQYRDLVYVSSGSDNITRLMD	429	
QY	267	POLKENSFG-----EGASDVLRLRVPMCDIWEIKFSCDHLHLSVAIGNDEGKYYWDL	320	
Db	430	IEC-----GACRLVLEGHELV-----RC-----IRFD---NKRIVSGAYGKIKVWDL	470	
QY	321	KSC---PPV---LITKLSHNOSKSVIRQTAMSDGSIILACDGGTIWRMDVI	367	
Db	471	QAALDPPRAPASTLCRLTVEH--SGRVER--LQDFEFOIISSHDITLIMDFL	520	
RESULT	5			
Y500_ANASP		STANDARD:	PRT:	1258 AA.
ID	Y500_ANASP			
AC	Q8YTC2;			
DT	01-MAR-2002 (Rel. 41, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical WD-repeat protein alr2800.			
GN	ALR2800.			
OS	Anabaena sp. (strain PCC 7120).			
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.			
OX	NCBI_TaxID=103690;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=21595285; PubMed=11759840;			
RA	Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,			
RA	Matanabe A., Iriuchl M., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kisida Y., Kohara M., Matsumoto M., Matsuo A., Muraki A.,			
RA	Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,			
RA	Yasuda M., Tabata S.;			
RT	"Complete genomic sequence of the filamentous nitrogen-fixing			
RT	cyanobacterium Anabaena sp. strain PCC 7120."			
RL	DNA Res. 8:205-213(2001).			
CC	-1- SIMILARITY: CONPAIRS 15 WD REPEATS (TRP-ASP DOMAINS).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboratio			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on it			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercia			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL, AP003590; BAB/4499.1;			
DR	PROSITE, PS00678; WD_REPEATS_1;			
DR	PROSITE, PS50082; WD_REPEATS_2;			
DR	PROSITE, PS50094; WD_REPEATS_REGION; 1.			
KM	Hypothetical protein: Repeat: WD repeat; Complete proteome.			
FT	REPEAT	55	93	WD 1.
FT	REPEAT	640	679	WD 2.
FT	REPEAT	682	721	WD 3.
FT	REPEAT	724	763	WD 4.
FT	REPEAT	766	807	WD 5.

FT REPEAT 809 849 WD 6.
 FT REPEAT 850 889 WD 7.
 FT REPEAT 892 931 WD 8.
 FT REPEAT 934 975 WD 9.
 FT REPEAT 976 1017 WD 10.
 FT REPEAT 1019 1059 WD 11.
 FT REPEAT 1060 1101 WD 12.
 FT REPEAT 1103 1143 WD 13.
 FT REPEAT 1144 1183 WD 14.
 FT REPEAT 1186 1227 WD 15.
 SO SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;

Query Match 9.8%; Score 193; DB 1; Length 1258;
 Best Local Similarity 23.2%; Pred. No. 3e-08;
 Matches 87; Conservative 56; Mismatches 134; Indels 98; Gaps 17;

QY 55 AGNRRITL-----YNCIDGAIASALQSYA-----DEDKESFYTVMACVNGN 98
 DB 588 AGGNVNLNLHNAQVDLSGDFSG---LTYWQAYLOGVNLHADVDFANSDLSCCVFTETLGN 644
 QY 99 PYVAGVKG-----IIRVIDVNSETIHKSLVGHGDSVNEIRPOLKPOLVITAS 148
 DB 645 ILTAAFEPEGQLATCDPDCDHVRWEVSKGLLLICRGHSNMVRFVFSR-DGEILLASCG 703
 QY 149 KDESRRLNWNVEGICILIFAGAGGHRVEYLSVDFHPSDIYRASCMDTTIKMSKEFW 208
 DB 704 ADENVKLSVNRGVCIKITLT--GHEHEVSVAFHP-DGETLASASGDKTIKMDIQD-G 758
 QY 209 FYVEKSFYTDPSKFEKFEVQFVFETASIHNYVDCNRMF--GDFILSKSVNELLME 266
 DB 759 TGLQ-----TLGHTDWRCAVAFSPDQNTLASADHTIKLMD 796
 QY 267 PQ-----LKNNSP-----GEGASDVLLR--YPPVPM-----DIW 293
 DB 797 VSQGRCLRTKSHGTGWRVSFAFADGQTLASGSDRTIKIMWYHGECEKTYIGHTNSY 856
 QY 294 FIKESCDLHLSVAIGNOGKRYVMDLKSCPVLITKLSHNSKSVIRRTANSVDSITL 353
 DB 857 STAYSPDSKI--LVSGSGGRITKLMW--CQTHICIKTLHGHTNEVC-SVAFSPDQTLA 910
 QY 354 ACCEDGTIWRMDVIT 368
 DB 911 CYSLDQSVRLMNCRT 925

RESULT 6
 ID T2D4_YEAST STANDARD: PRF: 798 AA.
 AC P38129;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Transcription initiation factor TFIID 90 kDa subunit (TAFII-90).
 GN TAF90 OR YBR198C OR YBR1410.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=95066384; PubMed=7975899;
 RA Mallet L., Bussiereau F., Jacquet M.;
 RT "Nucleotide sequence analysis of an 11.7 kb fragment of yeast
 RT chromosome II including BEM1, a new gene of the WD-40 repeat family
 RT and a new member of the KRE2/MNT1 family.";
 RL Yeast 10:819-831(1994).
 RN [2]
 RP SEQUENCE OF 726-747, AND CHARACTERIZATION.
 RC STRAIN=Y57;
 RX MEDLINE=95021683; PubMed=7935765;
 RA Reese J.C., Apone L., Walker S.S., Griffin L.A., Green M.R.;

RT "Yeast TAFIIS in a multisubunit complex required for activated
 RT transcription.";
 RL Nature 371:523-527(1994).
 RN [3]
 RP SEQUENCE OF 22-63 AND 726-752, AND CHARACTERIZATION.
 RC STRAIN=YPR252;
 RX MEDLINE=95396770; PubMed=7667272;
 RA Poon D., Bai Y., Campbell A.M., Bjorklund S., Kim Y.-J., Zhou S.,
 RA Kornberg R.D., Well P.A.;
 RT "Identification and characterization of a TFIID-like multiprotein
 RT complex from Saccharomyces cerevisiae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:8224-8228(1995).
 CC -1- FUNCTION: TAFs ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID
 CC (TFIID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF
 CC RNA POLYMERASE TRANSCRIPTION.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs) WHOSE MW RANGE FROM
 CC 25-150 kDa.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTRAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 CC -1- SIMILARITY: TO DROSOPHILA TAFII-80 AND MAMMALIAN TAF2D.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: 221487; CAA79685.1;
 DR EMBL: 236067; CAA85160.1;
 DR PIR: S34023; S34023.
 DR SGI: S0000402; TAF90.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.6.
 DR PRINTS: PR00320; GPROTEINBRT.
 DR SMART: SM00320; WD40.6.
 DR PROSITE: PS00678; WD_REPEATS_1; 3.
 DR PROSITE: PS50082; WD_REPEATS_2; 6.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Nuclear protein; Transcription regulation; Repeat; WD repeat.
 FT REPEAT 523 553 WD 1.
 FT REPEAT 565 595 WD 2.
 FT REPEAT 607 637 WD 3.
 FT REPEAT 649 679 WD 4.
 FT REPEAT 691 722 WD 5.
 SO SEQUENCE 798 AA; 88967 MW; B42315B8C752D0B6 CRC64;

Query Match 9.5%; Score 188; DB 1; Length 798;
 Best Local Similarity 20.4%; Pred. No. 4.3e-08;
 Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

QY 74 LOSYADEDEKESRYTYSMACGVNGNRYVAGGKGIIRVIDVNSETIH----- 121
 DB 459 MYTFQNTNMDSCGLDFSDDCRI-----AAAGFDSTYIKIWSLDSGLNPNPINALNNDK 513
 QY 122 -----KSLVGHGDSVNEIRPOLKPOLVITASKDESVRLNWNVEGICILIFAGAGHRE 176
 DB 514 EDPTCKTIVLGHSGTYVSTFSF--DNKYLLSGSDKTVRLMSDTHALVSYK---GHNP 569
 QY 177 VLSVDFHPSDIYRASCMDTTIKIWSKEFWTYVEKSFYTDPSKFEKFEVQFVFETA 236
 DB 570 VMDVFSFPLGHY-FATASHDQJATRLWSCDH-----YPLRI 605
 QY 237 SIHTNVVDCNRMF--GDFILSKSVNELLMEPOLKENSPEGASVLLRYPPKDI--- 292
 DB 606 AGLHNDVDCVSHPNNGCYVFTGSSDCTCRMDV-----STGDSVR-LFGLHAPVLSIV 659
 QY 293 -----WFIKSCDLHLSVAIGNOGKRYVMDLKSCPVLITKLSHNSKSVIRRTANSV 347
 DB 660 CPDGRV-----LSTGSEDCIINVMDIGTOKRL-----KQMGHCKNAIYLSYSYK 704

DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD_REPEATS_1; 3.
 DR PROSITE; PS50082; WD_REPEATS_2; 6.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repressor; Repeat; WD repeat.
 FT REPEAT 285 325 WD 1.
 FT REPEAT 332 371 WD 2.
 FT REPEAT 374 413 WD 3.
 FT REPEAT 415 454 WD 4.
 FT REPEAT 456 495 WD 5.
 FT REPEAT 510 549 WD 6.
 FT REPEAT 552 585 WD 7.
 SQ SEQUENCE 586 AA; 64192 MW; 79C991ABDDC7D095 CRC64;

Query Match 9.3%; Score 183.5; DB 1; Length 586;
 Best Local Similarity 22.1%; Pred. No. 6.9e-08;
 Matches 87; Conservative 62; Mismatches 161; Indels 83; Gaps 17;

QY 7 GNESTVGSITPSN-----KSKYKVTNRIOEGKKPLVAVNFPLD-----ARFF--DV 51
 DB 241 GNPPTVPAELIIPFSNVNREKDMVTSTNPNKEPPIVOLHLTEHTSVICVRFESADK 300
 QY 52 FYTAGGNRTILYNCIGDAISALQSYADEDEKESFYTVSWACGVNGNPVYAGVGKGIIR 111
 DB 301 FLATGCGNRAAMVENVETGKLTITLQESSKREGDLYRVSVAEPDCK-VLATGEEDQOIR 359
 QY 112 VIDVNSETHKSLVGHGDSVNEIRPOLKPOLYITASKDSVRLMNVETGICILIRAGAC 111
 DB 360 INDIAOKRVRRLTGHQOEIYSLDFSK-DGKTLVSGSGDRVCVLMVGEQKLIHLTD 418
 QY 172 GHRVEVLSVDFHPSDIYRASCMDTTIKMSKKEFTVVEKSFMTDPSKPTKPFYOF 231
 DB 419 G-----VTTFMFSPDQGF-IAAGSLDKVIRKWTSS--GTLVEQLHGHEES----- 460
 QY 232 PVEFTASIHNNYDCNRMFGDFILSKSVNDEILMEPOLKEN--SPG-----EGA 278
 DB 461 -YYSVAFSFD-----GKYLVSGLDNTIKLMELQCVSNVAPSKKEGICQKQTFGH 511
 QY 279 SDVLLRPPPMCDIWFITKISCDLHLSVAIGNQEGVYVWDLKSCPPVLTIKLSHNSOKS 338
 DB 512 KPEILSVTVSPDGKWTIS-----GSKRTIIOFSPSPS-PSHSLTGLGHNNSS-- 556
 QY 339 VIRQTAHSVDSITLACCEDGT-----IWRMD 365
 DB 557 -VISVAVSPNGH---CFATGSGDLRARIWSTE 584

RESULT 9
 PF20_CHLRE STANDARD; PRT; 606 AA.
 AC P93107;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Flagellar WD-repeat protein PF20.
 GN PF20.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadales; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith E., Lefevre P.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: INTER-MICROTUBULE BRIDGES IN FLAGELLA.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----

DR EMBL; U78547; AAB41727.1; -
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 7.
 DR PRINTS; PR00320; GPTEINBRPT.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS50082; WD_REPEATS_2; 6.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat; Flagella.
 FT REPEAT 324 354 WD 1.
 FT REPEAT 366 396 WD 2.
 FT REPEAT 408 438 WD 3.
 FT REPEAT 450 480 WD 4.
 FT REPEAT 492 522 WD 5.
 FT REPEAT 534 564 WD 6.
 FT REPEAT 576 606 WD 7.
 SQ SEQUENCE 606 AA; 65839 MW; 7A8779501E3218D1 CRC64;

Query Match 9.2%; Score 182; DB 1; Length 606;
 Best Local Similarity 23.5%; Pred. No. 9.6e-08;
 Matches 87; Conservative 57; Mismatches 134; Indels 92; Gaps 19;

QY 5 TLGNESTVGSITPSNKKSYKVTNRIOEGKKPLVAVNFPLDARFDFVYTAGGNRTILYN 64
 DB 236 SLGGQSTAAAGGAGSAGLCAITNRALTDVPPAA-----GAAAAAATGR----- 279
 QY 65 CLDGQATSAALQSYADEDEKESFYTVSWACGVNG-----NPVYAGVGKGIIRVIDVSETI 120
 DB 280 -----SGAVSAGP-----RSGNA-SLNAPRRNPADLEFPAPAKMLSLN----- 319
 QY 121 HKSIVGHGDSVNEIRPOLKPOLYITASKDSVRLMNVETGICILIFAGAGHRYEVLV 180
 DB 320 -KTFKGHLISVANILAHPTKP-ILVYASDDKTKMMHMPGG--DLIMCGE-GHKDWVAGV 374
 QY 181 DFHPSDIYRASCMDTTIKMSKKEFTVVEKSFMTDPSKPTKPFYOFPPFTAS 237
 DB 375 DFHPAGTC-LASGGDSAVKIMDF-----EKQRCVTTPTDHQO-----AIVSVR 417
 QY 238 IHTNNYDCNRMFGDFILSKSVNDEILMEPOLKENSPEGASVYLKYPVMCDIWFITK 297
 DB 418 FH-----HLGEVVASGSLDHTVRLMD-----LPAGCRNALRGHVDVNDLAKOPF 463
 QY 298 SCDLHLSVAIGNQEGVYVWDLKS--CPVLTIKLSHNSOKSVIROTAMSVDSITLAC 355
 DB 464 S-----SLATASSDKTVSWMDARAGLCIQ---TYGHQNSCGV--SFNIGTQLAST 512

QY 356 CEDGTIWRMD 365
 DB 513 DADGVVAKLMD 522

RESULT 10
 GBB_MAIZE STANDARD; PRT; 380 AA.
 ID GBB_MAIZE
 AC P49178;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Guanine nucleotide-binding protein beta subunit.
 GN GBI.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Lillipsidea; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95023950; PubMed=7937804;
 RA Weiss C.A., Garnaat C.W., Mukai K., Hu Y., Ma H.;

```

RT      * Isolation of cDNAs encoding guanine nucleotide-binding protein beta-
RT      subunit homologues from maize (ZGB1) and Arabidopsis (AGB1).".
RL      Proc. Natl. Acad. Sci. U.S.A. 91:9554-9558(1994).
CC      -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC      INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
CC      SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
CC      GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
CC      EFFECTOR INTERACTION.
CC      -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC      -1- TISSUE SPECIFICITY: PRESENT IN THE ROOT, LEAF, AND TASSEL.
CC      -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U12233; AAA50446.1; -.
CC      DR      HSSP: P04901; 1TBBG.
CC      DR      MalzEDB: 105669; -.
CC      DR      InterPro: IPR001632; Gprotein_B.
CC      DR      InterPro: IPR001680; WD40.
CC      DR      Pfam: PF00400; WD40; 7.
CC      DR      PRINTS: PR00320; GPROTEINB.
CC      DR      SMART: SM00320; WD40; 7.
CC      DR      PROSITE: PS00678; WD_REPEATS_1; 3.
CC      DR      PROSITE: PS50082; WD_REPEATS_2; 5.
CC      DR      PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC      KW      Transducer; Repeat; WD repeat.
CC      FT      REPEAT 64 94 WD 1.
CC      FT      REPEAT 106 136 WD 2.
CC      FT      REPEAT 155 186 WD 3.
CC      FT      REPEAT 203 234 WD 4.
CC      FT      REPEAT 247 277 WD 5.
CC      FT      REPEAT 296 326 WD 6.
CC      FT      REPEAT 342 372 WD 7.
CC      FT      REPEAT 380 AA; 41714 MW; 7D5DC5276C3DD1E CRC64;
CC      SQ      SEQUENCE 380 AA; 41714 MW; 7D5DC5276C3DD1E CRC64;

Query Match          9.2%; Score 181.5; DB 1; Length 380;
Best Local Similarity 24.9%; Pred. No. 5,7e-08;
Matches 80; Conservative 46; Mismatches 142; Indels 53; Gaps 12;

OY      59 RITLYNCLGDAISALOSYADEDEKESFYYSMACGVNPNFYVAAGYKGIIRIVDNSE 118
DB      89 RLTYWNAITLSQKTHAIKIHCP-----WVWACAFAPNGOS-VACGGIDSAISIFNLNSQ 140
OY      119 T-----IHSLYGHGHSVNEIRTPQIKPOLVITASKDESVRLMNVETGICLIIFAG- 169
DB      141 ADRGNMNPVSHLTHGHKVSVCQYVPDQETRLITSSSDQCVLMVDTTGORISIFGGEF 200
OY      170 AGHRYEVLSDYFHPSDIYFASCGMDTITIKWSMKEFWYVEKSFYWTDDPSKPTFEV 229
DB      201 PSGRHADVQSYSSINSMFMVSGSCDTYVLMDRISRAVRYTHGHEDVNS--VAF- 257
OY      230 QPVTFTASHTNYVDNFMFGDFILSKSVNDELIMEP-----OLKNSPEGASDVLL 283
DB      258 -FP-----DGNH-FG-----TGSDDGTCRLFDKRTGHQLOVYSREDDRNSE--- 297
OY      284 RYVPMPDMCIWIKFSCDHLHSVAAGNEGKYVWDLSKSPVLLITKLSHNSKSVINQI 343
DB      298 -----LPTVTSIAFSISGRLL--LFAGYSGDCQYVDTLLAEVVLNGLNLSHDSRISCL 350
OY      344 AMSVDSFTILACCEDEGTIRW 364
DB      351 GMSSDGSAICTGSMDKNLKIW 371
RESULT 11

```

```

TRCB_XENLA
ID      TRCB_XENLA          STANDARD;          PRT;          518 AA.
AC      Q91854; P70037; P70038;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Beta-Trcp (beta-transducin repeat-containing protein).
GN      FBXW1 OR BTCP.
OS      Xenopus laevis (African clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC      Xenopodidae; Xenopus.
OX      NCBI_TaxID=8353;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9330289; PubMed=8393141;
RA      Szelek W., Keiper B.D., Stratowa C., Castanon M.J.;
RT      "Saccharomyes cerevisiae cdc15 mutants arrested at a late stage in
RT      anaphase are rescued by Xenopus cDNAs encoding N-ras or a protein
RT      with beta-transducin repeats."
RT      Mol. Cell. Biol. 13:4953-4966(1993).
RN      [2]
RP      SEQUENCE OF 302-518 FROM N.A.
RX      MEDLINE=97109804; PubMed=8952061;
RA      Hudson J.W., Alarcon V.B., Elinson R.P.;
RT      "Identification of new localized RNAs in the Xenopus oocyte by
RT      differential display PCR."
RL      Dev. Genet. 19:190-198(1996).
CC      -1- FUNCTION: Probably recognizes and binds to some phosphorylated
CC      proteins and promotes their ubiquitination and degradation.
CC      -1- SUBUNIT: PART OF A SCF (SKP1-CUL1IN-F-BOX) PROTEIN LIGASE COMPLEX
CC      (BY SIMILARITY).
CC      -1- DEVELOPMENTAL STAGE: PRESENT IN FULLY GROWN AND PROGESTERONE-
CC      MATURED OOCYTES. THE LEVEL CHANGE VERY LITTLE EVEN AFTER ZYGOTIC
CC      GENE TRANSCRIPTION BEGINS FOLLOWING THE MIDBLASTULA TRANSITION. DO
CC      NOT INCREASE IN ABUNDANCE IN THE GASTRULA, NEURULA, TAILBUD, OR
CC      TADPOLE EMBRYO.
CC      -1- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.
CC      -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC      -1- SIMILARITY: STRONG, TO C.ELEGANS K10B2.1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M96268; AAA02810.1; -.
CC      DR      EMBL: U63921; AAB49671.1; -.
CC      DR      EMBL: U63922; AAB49672.1; -.
CC      DR      InterPro: IPR001810; F-box.
CC      DR      InterPro: IPR001680; WD40.
CC      DR      Pfam: PF00646; F-box; 1.
CC      DR      Pfam: PF00400; WD40; 7.
CC      DR      PRINTS: PR00320; GPROTEINBPT.
CC      DR      SMART: SM00256; FBOX; 1.
CC      DR      SMART: SM00320; WD40; 7.
CC      DR      PROSITE: PS50181; FBXW1; 1.
CC      DR      PROSITE: PS00678; WD_REPEATS_1; 6.
CC      DR      PROSITE: PS50082; WD_REPEATS_2; 7.
CC      DR      PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC      KW      Ubiquitin conjugation; Repeat; WD repeat.
CC      FT      DOMAIN 119 157 F-BOX.
CC      FT      REPEAT 230 258 WD 1.
CC      FT      REPEAT 270 298 WD 2.
CC      FT      REPEAT 310 338 WD 3.
CC      FT      REPEAT 353 381 WD 4.
CC      FT      REPEAT 393 421 WD 5.
CC      FT      REPEAT 433 461 WD 6.
CC      FT      REPEAT 482 510 WD 7.
CC      FT      REPEAT 502 530 GEM -> EFR (IN REF. 2).

```

FT CONFLICT 516 518 GIA -> AAH (IN REF. 2).
 SQ SEQUENCE 518 AA: 59507 MW: 2A52EC19028127F3 CRC64;

Query Match 9.1%; Score 179.5; DB 1; Length 518;
 Best Local Similarity 25.4%; Pred. No. 1.3e-07;
 Matches 75; Conservative 45; Mismatches 108; Indels 67; Gaps 14;

QY 87 YVSNACGVNAGNPNVAAAGVKGIIIRYIDVNSFIHKSIVGHGSDVNEIRTPQLPOLVIT 146
 DB 271 HTGVSVCLODYDERVITITGSSDSTVRWVDVNTGEMLTLIHCEAVLHLR--FNNGMAMV 327
 QY 147 ASKDESRLMNVETGICILIFAGAGHRYEVLSDPHSPDIYFASCGMDTTIKISMKE 206
 DB 328 CSKDRSIAMVDMASTPDIITLRLVYGHRAAVNVDF--DDKIYVSAAG--DRTIKVMN--- 381
 QY 207 FMTYVEKSTFTWDDPSKFPKFPVFTASITHTNVDCNRMGDFILSKSVNELLIME 266
 DB 382 -----TSTCEF-VRTINGHKGRIACIQYDRLVVSGSSDNTIRLMD 421
 QY 267 POLKENSFG-----EGASDVLRLRYVPMCDIWFIFSCDLHSSVAIGNQEGKYVWML 320
 DB 422 IEC-----GACLRVLEGHELV-----RC-----IRFD-----NKRIVSAGAYDKIKIYVMDL 462
 QY 321 KSC-----PPV-----LITKLSHNSKSVIRQTAMSVDSSTLIACCEGDTIRMDVI 367
 DB 463 VALDPRAPAGTICLRLTYEH--SGRVFR---LQDFEQLVSSSHDITILIMDEL 512

RESULT 12

FWLA_HUMAN STANDARD: PRT: 605 AA.

AC 09Y297: 09Y213: 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE F-box/WD-repeat protein 1B (F-box and WD-repeats protein beta-TrCP)
 DE (E3SISIKAPAB) (pikappaalpha-E3 receptor subunit).
 GN FBXK1A OR FBX1A OR BTRCP OR BTRC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=99075339; PubMed=9859996;
 RA Yaron A., Hatzubai A., Davis M., Lavon I., Amit S., Manning A.M.,
 RA Andersen J.S., Mann M., Mercurio F., Ben-Neriah Y.;
 RT "Identification of the receptor component of the Ikapalpha-
 RT ubiquitin ligase.";
 RL Nature 396:590-594(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Lymphoid;
 RX MEDLINE=98325370; PubMed=9660940;
 RA Margottin F., Bour S.P., Durand H., Selig L., Benichou S., Richard V.,
 RA Thomas D., Strebel K., Benatou R.;
 RT "A novel human WD protein, h-beta trcp, that interacts with HIV-1 Ypu
 RT connects CD4 to the ER degradation pathway through an F-box motif.";
 RL Mol. Cell 1:565-574(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=20003060; PubMed=10531035;
 RA Centaurelli C., Chlaur D.S., Guadavaccaro D., Parks W., Vidal M.,
 RA Pagano M.;
 RT "Identification of a family of human F-box proteins.";
 RL Curr. Biol. 9:1177-1179(1999).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=99145464; PubMed=9990852;
 RA Winston J.T., Strack P., Beer-Romero P., Chu C.Y., Elledge S.J.,
 RA Harper J.W.;
 RT "The SCF(beta-TRCP)-ubiquitin ligase complex associates specifically

RT with phosphorylated destruction motifs in I-kappa-B-alpha and
 RT beta-catenin and stimulates I-kappa-B-alpha ubiquitination in vitro.";
 RL Genes Dev. 13:270-283(1999)
 CC - FUNCTION: BINDS SPECIFICALLY TO PHOSPHORYLATED IKB ALPHA
 CC (PIKAPABALPHA) AND PHOSPHORYLATED BETA-CATENIN AND PROMOTES THEIR
 CC UBIQUITINATION AND DEGRADATION.
 CC - SUBUNIT: PART OF A SCF (SKP1-CUL1N-F-BOX) PROTEIN LIGASE COMPLEX.
 CC - SUBCELLULAR LOCATION: Cytoplasmic.
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC - SIMILARITY: CONTAINS 1 F-BOX DOMAIN.
 CC - SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AF101784; AAD08702.1; -
 DR EMBL: YL4153; CAA14572.1; -
 DR EMBL: AF129530; AAF04464.1; -
 DR MIM: 603482; -
 DR InterPro: IPR001810; F-box.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00646; F-box; 1.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PR00320; GPROTEINBPT.
 DR SMART: SM00256; FBOX; 1.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS50181; FBOX; 1.
 DR PROSITE: PS500678; WD_REPEATS_1; 6.
 DR PROSITE: PS50082; WD_REPEATS_2; 7.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KM Ubiquitin conjugation: Repeat; WD repeat; Alternative splicing.
 FT DOMAIN 190 228 F-BOX.
 FT REPEAT 301 338 WD 1.
 FT REPEAT 341 378 WD 2.
 FT REPEAT 381 418 WD 3.
 FT REPEAT 424 461 WD 4.
 FT REPEAT 464 503 WD 5.
 FT REPEAT 505 541 WD 6.
 FT REPEAT 553 590 WD 7.
 FT VARSPLIC 17 52 MISSING (IN ISOFORM 2).
 SQ SEQUENCE 605 AA: 68866 MW: 4C67F3B7E400ED37 CRC64;

Query Match 9.0%; Score 177.5; DB 1; Length 605;
 Best Local Similarity 25.4%; Pred. No. 2.3e-07;
 Matches 75; Conservative 45; Mismatches 108; Indels 67; Gaps 14;

QY 87 YVSNACGVNAGNPNVAAAGVKGIIIRYIDVNSFIHKSIVGHGSDVNEIRTPQLPOLVIT 146
 DB 342 HTGVSVCLODYDERVITITGSSDSTVRWVDVNTGEMLTLIHCEAVLHLR--FNNGMAMV 398
 QY 147 ASKDESRLMNVETGICILIFAGAGHRYEVLSDPHSPDIYFASCGMDTTIKISMKE 206
 DB 399 CSKDRSIAMVDMASTPDIITLRLVYGHRAAVNVDF--DDKIYVSAAG--DRTIKVMN--- 452
 QY 207 FMTYVEKSTFTWDDPSKFPKFPVFTASITHTNVDCNRMGDFILSKSVNELLIME 266
 DB 453 -----TSTCEF-VRTINGHKGRIACIQYDRLVVSGSSDNTIRLMD 492
 QY 267 POLKENSFG-----EGASDVLRLRYVPMCDIWFIFSCDLHSSVAIGNQEGKYVWML 320
 DB 493 IEC-----GACLRVLEGHELV-----RC-----IRFD-----NKRIVSAGAYDKIKIYVMDL 533
 QY 321 KSC-----PPV-----LITKLSHNSKSVIRQTAMSVDSSTLIACCEGDTIRMDVI 367
 DB 534 VALDPRAPAGTICLRLTYEH--SGRVFR---LQDFEQLVSSSHDITILIMDEL 583

```

RESULT 13
GBL_SOLUTION
ID GBL_SOLUTION STANDARD: PRT; 377 AA.
AC P93563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit.
GN GBL.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. DESIREE;
RA Provart N.J., Ma H., Willmitzer L., Mueller-Roeber B.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEIN (G PROTEIN) ARE
CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
CC EFFECTOR INTERACTION.
CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X87837; CAA61106.1; -
CC HSSP; P04901; 178G.
CC InterPro: IPR001632; Gprotein_B.
CC InterPro: IPR001680; WD40.
CC Pfam; PF00400; WD40; 7.
CC PRINTS; PR00319; GPROTEINB.
CC SMART; SM00320; GPROTEINBPT.
CC PROSITE; PS00678; WD_REPEATS_1; 2.
CC PROSITE; PS50082; WD_REPEATS_2; 5.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Transducer; Repeat; WD repeat.
CC FT REPEAT 63 93 WD 1.
CC FT REPEAT 105 135 WD 2.
CC FT REPEAT 154 185 WD 3.
CC FT REPEAT 202 233 WD 4.
CC FT REPEAT 246 276 WD 5.
CC FT REPEAT 293 323 WD 6.
CC FT REPEAT 339 369 WD 7.
CC SEQUENCE 377 AA; 40941 MW; B9AFLAAGT2IDCBCL CRC64;

Query Match 8.9%; Score 176; DB 1; Length 377;
Best Local Similarity 23.2%; Pred. No. 1.6e-07;
Matches 86; Conservative 49; Mismatches 156; Indels 80; Gaps 14;
QY 30 OEGKKPLAVVFNFLD-----ARFDFVTAGNRRTLYNCLGDAISALQSYA 78
DB 42 RQKSP---VFEGPTDLVCCRLGHTGKYSIDMTPEKNRIV--SASODGLIYWNAL 96
QY 79 DEKKESEFYTVSWACGVGNP---VVAAGVGGIIRVIDVNS---ETIH---KSLVGHG 128
DB 97 SQKTAHAIKPCAMWMTCAFSPSGGVACGGLDSACSIFNLNPSIDKGIHPSRMLSGHK 156
QY 129 DSVNEIRIOPPLKPOLVITASKDESRLKNNVETGICILIFAG--AGCHREVLVSYPFHSD 186
DB 157 GYVSSCQYVPEDTHLITSSGDQCVLMDITTGRTSVFGGEFGHGTADVLVSISSEN 216

```

```

RESULT 14
PKWA_THECU
ID PKWA_THECU STANDARD: PRT; 742 AA.
AC P49695;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative serine/threonine-protein kinase pkwa (EC 2.7.1.-).
GN PKWA OR PKW1.
OS Thermomonospora curvata.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
OC Thermomonospora.
OX NCBI_TaxID=2020;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CCM 3352;
RX MEDLINE=96200125; PubMed=8631732;
RA Janda L., Richy P., Spizak J., Petricek M.;
RT "A deduced Thermomonospora curvata protein containing
RT serine/threonine protein kinase and WD-repeat domains."
RL J. Bacteriol. 178:1487-1489(1996).
CC -1- FUNCTION: MAY PLAY A REGULATORY ROLE DURING THE COMPLEX GROWTH
CC CYCLE AND IN SECONDARY METABOLITE PRODUCTION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF115313; AAB05822.1; -
CC HSSP; P00523; 2PTK
CC InterPro: IPR000719; Euk_Pkinase.
CC InterPro: IPR002290; Ser_thr_Pkinase.
CC InterPro: IPR001680; WD40.
CC Pfam; PF00069; Pkinase; 1.
CC Pfam; PF00400; WD40; 7.
CC PRINTS; PR00320; GPROTEINBPT.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS00678; WD_REPEATS_1; 5.
CC PROSITE; PS00682; WD_REPEATS_2; 7.
CC PROSITE; PS50082; WD_REPEATS_REGION; 1.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Transducer; Serine/threonine-protein kinase; ATP-binding; Repeat;
CC WD repeat.
CC KW DOMAIN 16 266 PROTEIN KINASE.
CC NP_BIND 22 30 ATP (BY SIMILARITY).
CC FT BINDING 44 44 ATP (BY SIMILARITY).

```

```

CC -1- COFACTOR: MAGNESIUM OR MANGANESE.
CC -1- SUBUNIT: OLIGOMER.
CC -1- DOMAIN: CONSISTS OF AN N-TERMINAL DOMAIN WITH PROBABLE COILED COIL
CC STRUCTURE, A CENTRAL NONREPEATITIVE CATALYTIC DOMAIN, AND A C-
CC TERMINAL DOMAIN WITH SEVEN WD REPEATS.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE MHCK / EF-2 PROTEIN KINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U16856; AAA6670.1; -.
DR DctyDb: DD01086; mhka.
DR InterPro: IPR004166; MHCK_EF2_kinase.
DR InterPro: IPR001680; WD40.
DR Pfam: PF02816; MHCK_EF2_kinase; 1.
DR Pfam: PF00400; WD40; 7.
DR PRINTS: PR00320; GPROTEINRPT.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS50082; WD_REPEATS_2; 5.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Transferrase: Serine/threonine-protein kinase; ATP-binding; Repeat;
KW WD repeat; Phosphorylation; Coiled coil.
FT DOMAIN 100 120 Coiled coil (POTENTIAL).
FT DOMAIN 144 148 POLY-GLN.
FT DOMAIN 175 181 POLY-GLY.
FT DOMAIN 187 241 COILED COIL (POTENTIAL).
FT DOMAIN 297 502 COILED COIL (POTENTIAL).
FT DOMAIN 345 348 POLY-SER.
FT DOMAIN 438 441 POLY-LEU.
FT DOMAIN 500 551 PSEUDOSUBSTRATE/AUTOINHIBITORY DOMAIN
(POTENTIAL).
FT NP_BIND 552 852 852 783 ATP (POTENTIAL).
FT REPEAT 867 897 897 897 WD 1.
FT REPEAT 910 938 938 938 WD 2.
FT REPEAT 952 980 980 980 WD 3.
FT REPEAT 993 1021 1021 1021 WD 4.
FT REPEAT 1033 1061 1061 1061 WD 5.
FT REPEAT 1073 1101 1101 1101 WD 6.
FT REPEAT 1114 1142 1142 1142 WD 7.
FT REPEAT 1146 1146 1146 1146 WD 8.
SQ SEQUENCE 1146 AA; 128945 MW; 980831F7948B573 CRC64;

Query Match 8.9%; Score 175.5; DB 1; Length 1146;
Best Local Similarity 22.4%; Pred. No. 8e-07;
Matches 89; Conservative 55; Mismatches 138; Indels 115; Gaps 20;

QY 5 TLGNESIVSLTPSNKKSKYVTRNIOECKKPLAYVNFELDARFPDVEVTAGNRIITYN 64
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
Db 825 TMVMPDILPELMPD-----NTIKVAGKQPKRAEFSKDKL----- 860

QY 65 CLIGDAISALDSYADEDEKEESFTYYSMACGVNGNPFYVAAGYKGLIRYIDVNSE---T 119
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
Db 861 -----CVSTIOSFRERVSIAFF-----DNOKILCAGYDGTFRVFDVDMNKCLYT 907

QY 120 IHKSLVAGDSVNEIRTPOLKPOLVITASKDESVRLMWVNGS-ICILIFAGGGRREV 177
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
Db 908 VN-----GHRKSIEST---ACNSNTYFTSSPDNTIVHILIRSNNTYCIETLV---GHTGEV 957

QY 178 LSYDFHPSDIYRFASCGMDTITIKISMKEFWTYVE-KSFTWTDPSKEPTKVPDPVFTA 236
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
Db 958 NCV--VNAEKLTF-SCSYDKTIKIVDLS---TFKEIKSP-----E 991

QY 237 SIHTNYVDCNRWFGDFILSKSYNEIILMEPO-----LKENSPGE-----GASDVILR 284
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

```

```

Db      992  GVHTRKIKTALSGRYLFSGSGNDQIIYVWDTFLSLMFNMQGHBMVLSLHCTASYLEST 1051
OY      285  YPVPMDQIIML--RFS- ----DLHLSVA-----IGNEGRVVYWDLKSQCPVLT 329
Db      1052 SKDVIKTIWDLNSFCILDTLKGHMNSVSCVYKDRILYSGSBDNSIKYWDLDLTLEYVTI 1111
OY      330  KLSHNQSKSVYRQTAMSVDSQSTLLACCCEGT--RHW 364
Db      1112 PKSHS-----LVGKCLMVFENNQILISAFAFDSIKVWEM 1143

```

Search completed: June 25, 2002, 01:25:37
Job time: 621 sec